Максимально правдоподобно о моделях бинарного выбора

Эконометрика. Лекция 7

Метод максимального правдоподобия

Наблюдения: вижу работающий фонтан
Гипотеза 1: фонтан работает каждый день
Гипотеза 2: фонтан включают раз в году

При какой гипотезе вероятность имеющихся данных максимальна?

Правдоподобие. Более формально

Метод максимального правдоподобия (ML — Maximum Likelihood)
В качестве оценки неизвестного параметра θ возьмем такое число $\hat{\theta}$, при котором вероятность имеющихся данных максимальна.

Задача. Дискретный случай [у доски]

Наблюдения: $y_1 = 0$, $y_2 = 1$, $y_3 = 2$, $y_4 = 0$.
Модель: наблюдения y_i независимы,

y_i	0	1	2
Вероятность	p	$2p$	$1 - 3p$

Найдите $\hat{p}$ с помощью метода максимального правдоподобия

Правдоподобие. Непрерывный случай

Для непрерывных случайных величин максимизируется плотность вероятности

Для независимых наблюдений:

$f(y_1, y_2, ..., y_n|\theta) = f(y_1|\theta) \cdot f(y_2|\theta) \cdot ... \cdot f(y_n|\theta) = \prod f(y_i|\theta)$

Трюк с логарифмированием: $\ell(\theta) = \ln\left(\prod f(y_i|\theta)\right) = \sum \ln f(y_i|\theta)$

Задача. Непрерывный случай [у доски]

100 наблюдений: $y_1 = 1.1$, $y_2 = 2.7$, …, $y_{100} = 1.5$.
Сумма, $\sum y_i = 200$.
Модель: наблюдения независимы, $f(y) = \lambda e^{-\lambda x}$ при $x > 0$.
Найдите $\hat{\lambda}$ с помощью метода максимального правдоподобия

ML — это хорошо!

Оценка $\hat{\theta}_{ML}$ — случайная величина
ML оценки:

- Состоятельны: $\hat{\theta}_{ML} \to \theta$ при $n \to \infty$
- Асимптотически несмещены: $E(\hat{\theta}_{ML}) \to \theta$ при $n \to \infty$
- Асимптотически эффективны:

$Var(\hat{\theta}_{ML})$ наименьшая среди асимптотически несмещенных оценок

ML — это нормально!

- Асимптотически нормальны:

$\hat{\theta}_{ML} \sim N(\theta, I^{-1})$ при $n >> 0$
I — информация Фишера, $I = -E\left(\ell''(\theta)\right)$
В многомерном случае: $I = -E(H)$, H — матрица Гессе

ML оценка как случайная величина

Среднее: $E(\hat{\theta}_{ML}) \approx \theta$, дисперсия: $Var(\hat{\theta}_{ML}) \approx I^{-1}$
Оценка дисперсии: $\widehat{Var}(\hat{\theta}_{ML}) = \hat{I}^{-1}$
Наблюдаемая информация Фишера $\hat{I} = -\ell''(\hat{\theta})$

Доверительный интервал

Доверительный интервал:

$$\theta \in [\hat{\theta}_{ML} - z_{cr} se(\hat{\theta}); \hat{\theta}_{ML} + z_{cr} se(\hat{\theta})],$$

$se(\hat{\theta}) = \sqrt{\widehat{Var}(\hat{\theta}_{ML})} = \sqrt{-(l''(\hat{\theta}))^{-1}}$

Продолжение задачи [у доски]

100 наблюдений: $y_1 = 1.1$, $y_2 = 2.7$, ..., $y_{100} = 1.5$.
Сумма, $\sum y_i = 200$.
Модель: наблюдения независимы, $f(y) = \lambda e^{-\lambda x}$ при $x > 0$.
Постройте 95%-ый доверительный интервал для λ.

Проверка гипотез

H_0: Система из q уравнений на неизвестные параметры
H_a: Хотя бы одно из q условий не выполнено
Тест отношения правдоподобия (Likelihood Ratio, LR):

$$LR = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_{H_0})) \sim \chi^2_q$$

Продолжение задачи [у доски]

100 наблюдений: $y_1 = 1.1$, $y_2 = 2.7$, …, $y_{100} = 1.5$.
Сумма, $\sum y_i = 200$.
Модель: наблюдения независимы, $f(y) = \lambda e^{-\lambda x}$ при $x > 0$.
Проверьте гипотезу H_0: $\lambda = 1$.

Логит и пробит-модели

Бинарная объясняемая переменная: $y_i \in \{0, 1\}$.
Скрытая ненаблюдаемая переменная: $y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i$.

$$y_i = \begin{cases} 1, y_i^* \geq 0 \\ 0, y_i^* < 0 \end{cases}$$

Разница логит-пробит

- Пробит-модель: $\varepsilon_i \sim N(0, 1)$.
- Логит-модель: $\varepsilon_i \sim logistic$, $f(t) = e^{-x}/(1 + e^{-x})^2$
- Логистическое распределение похоже на нормальное $N(0, 1.6^2)$

Вероятность

$$
\begin{aligned}
P(y_i = 1) = P(y_i^* \geq 0) = P(\beta_1 + \beta_2 x_i + \varepsilon_i \geq 0) = \\
= P(-\varepsilon_i \leq \beta_1 + \beta_2 x_i) = P(\varepsilon_i \leq \beta_1 + \beta_2 x_i) = \\
= F(\beta_1 + \beta_2 x_i)
\end{aligned}
$$

Упражнение [у доски]

Для логит-модели найдите $P(y_i = 1)$, $\ln P(y_i = 1)/P(y_i = 0)$

Логарифмическое отношение шансов

Для логит-модели:

- Вероятность:

$$P(y_i = 1) = \frac{exp(\beta_1 + \beta_2 x_i)}{1 + exp(\beta_1 + \beta_2 x_i)} = \frac{1}{1 + exp(-(\beta_1 + \beta_2 x_i))}$$

- Логарифмическое отношение шансов:

$$\ln \frac{P(y_i = 1)}{P(y_i = 0)} = \beta_1 + \beta_2 x_i$$

Функция правдоподобия

Наблюдения: $y_1 = 1$, $y_2 = 0$, $y_3 = 1$, …
Модель: логит.
Функция правдоподобия:

$$P(y_1 = 1, y_2 = 0, ...) = P(y_1 = 1) \cdot P(y_2 = 0) \cdot P(y_3 = 1) \cdot ...$$

Вероятность и отношение шансов [у доски]

Интерпретация

Коэффициенты плохо интерпретируемы
Предельный эффект — производная вероятности:

$$\frac{dP(y=1)}{dx} = \frac{dF(\beta_1 + \beta_2 x)}{dx} = \beta_2 \cdot f(\beta_1 + \beta_2 x)$$

Зависит от x (!)

Два средних предельных эффекта:

Средний предельный эффект по наблюдениям:

$$\frac{\sum \beta_2 \cdot f(\beta_1 + \beta_2 x_i)}{n}$$

Предельный эффект для среднего наблюдения:

$$\beta_2 \cdot f(\beta_1 + \beta_2 \bar{x})$$

Прогнозирование

- Прогноз скрытой переменной: $\hat{y}_f^* = \hat{\beta}_1 + \hat{\beta}_2 x_f$
- Точечный прогноз вероятности: $\hat{P}(y_f = 1) = F(\hat{y}_f^*)$
- Доверительный интервал для $E(\hat{y}_f^*)$:

$$[\hat{y}_f^* - z_{cr} se(\hat{y}_f^*); \hat{y}_f^* + z_{cr} se(\hat{y}_f^*)]$$

- Доверительный интервал для вероятности:

$$[F(\hat{y}_f^* - z_{cr} se(\hat{y}_f^*)); F(\hat{y}_f^* + z_{cr} se(\hat{y}_f^*))]$$

- В R: логистическая функция распределения, $F(.) = plogis(.)$, нормальная $F(.) = pnorm(.)$.

Разница логит-пробит на практике

Коэффициенты логит/пробит на практике отличаются в ~ 1.6 раза:

- Логит-модель: $y_i^* = \beta_1 + \beta_2 x_i + u_i$, где u_i примерно $N(0, 1.6^2)$
- Логит-модель: $\frac{y_i^*}{1.6} = \frac{\beta_1}{1.6} + \frac{\beta_2}{1.6} x_i + \frac{u_i}{1.6}$, где $\frac{u_i}{1.6}$ примерно $N(0, 1)$
- Пробит-модель: $y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i$, где $\varepsilon_i \sim N(0, 1)$
- $\{y_i = 1\} \Leftrightarrow \{y_i^* > 0\} \Leftrightarrow \{y_i^*/1.6 > 0\}$

Проблема логит-пробит моделей [у доски]

“Идеальное прогнозирование”:

$y_1 = 0$	$y_2 = 0$	$y_3 = 1$
$x_1 = 1$	$x_2 = 2$	$x_3 = 3$

ML оценки не существуют!

Проблема логит-пробит моделей

- Нередко возникает при большом количестве дамми-регрессоров
- Признаки: не сходится ML,

R: “fitted probabilities numerically 0 or 1 occurred”

- Решения: регуляризация (добавление штрафа к $\ell()$), байесовский подход

Мораль

- Метод максимального правдоподобия. Позволяет получать оценки неизвестных параметров.
- Логит и пробит модели. Модели для зависимой переменой, принимающей значения 0 и 1.
- МНК не подходит для моделирования бинарной зависимой переменной

Источники мудрости:

- Борзых Д.А., Демешев Б.Б. Эконометрика в задачах и упражнениях: глава 5
- Катышев П.К., Пересецкий А. А. Эконометрика. Начальный курс: главы 10.1, 10.2, 12.1
- Носко В.П., Эконометрика для начинающих, дополнительные главы (www.iep.ru/files/persona/nosko/Book.pdf) : главы 1.1 – 1.4