

Лекция 7. Максимально правдоподобно о логит и пробит-моделях

8 октября 2014 г.

Метод максимального правдоподобия

Наблюдения: вижу работающий фонтан

Гипотеза 1: фонтан работает каждый день

Гипотеза 2: фонтан включают раз в году

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.

Пример задачи

Наблюдения: $y_1 = 0$, $y_2 = 1$, $y_3 = 2$, $y_4 = 0$.

Модель: наблюдения независимы,

y	0	1	2
Вероятность	p	$2p$	$1 - 3p$

Решаем задачу у чудо-доски

Правдоподобие. Непрерывный случай

Для непрерывных случайных величин максимизируется плотность вероятности

Для независимых наблюдений:

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

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

Задача 2.

100 наблюдений: $y_1 = 1.1$, $y_2 = 2.7$, \dots , $y_{100} = 1.5$.

Сумма, $\sum y_i = 200$.

Модель: наблюдения независимы, $f(y) = \lambda e^{-\lambda x}$ при $x > 0$.

Найдите $\hat{\lambda}$

Решаем задачу 2 чудо-доска.

ML — это хорошо!

ML оценки:

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

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

ML — это нормально!

- ▶ Асимптотически нормальны:

$$\hat{\theta}_{ML} \sim N(\theta, I^{-1}) \text{ при } n \gg 0$$

I — информация Фишера, $I = -E(I''(\theta))$

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 $\hat{I} = -l''(\hat{\theta})$

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 чудо-доски

Постройте 95%-ый доверительный интервал для θ .

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):

$$LR = 2(l(\hat{\theta}) - l(\hat{\theta}_{H_0})) \sim \chi_q^2$$

Продолжение задачи у чудо-доски

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

Логит-модель: $\varepsilon_i \sim \text{logistic}$, $f(t) = e^{-x}/(1 + e^{-x})^2$

Пробит-модель: $\varepsilon_i \sim N(0, 1)$.

Логистическое похоже на $N(0, 1.6^2)$

$$\begin{aligned} P(y_i = 1) &= P(y_i^* \geq 0) = P(\beta_1 + \beta_2 x_i + \varepsilon_i \geq 0) = \\ &= P(-\varepsilon_i \leq \beta_1 + \beta_2 x_i) = P(\varepsilon_i \leq \beta_1 + \beta_2 x_i) = \\ &= F(\beta_1 + \beta_2 x_i) = \int_{-\infty}^{\beta_1 + \beta_2 x_i} f(t) dt \quad (1) \end{aligned}$$

Упражнение.

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

Чудо-Доска

Логарифмическое отношение шансов

Для логит-модели:

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

$$\ln P(y_i = 1)/P(y_i = 0) = \beta_1 + \beta_2 x_i$$

Функция правдоподобия

Наблюдения: $y_1 = 1, y_2 = 0, \dots$

Модель: логит.

Функция правдоподобия:

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

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: $\hat{y}_f^* = \hat{\beta}_1 + \hat{\beta}_2 x_f$

Доверительный интервал для $E(\hat{y}_f^*)$:

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

Переход к $\hat{P}(y_f = 1) = F(y_f^*)$

Разница логит-пробит на практике

Коэффициенты логит/пробит отличаются в ~ 1.6 раза:

Логит (примерно): $y_i^* = \beta_1 + \beta_2 x_i + N(0, 1.6^2)$

$$\frac{y_i^*}{1.6} = \frac{\beta_1}{1.6} + \frac{\beta_2}{1.6} x_i + N(0, 1)$$

Пробит: $y_i^* = \beta_1 + \beta_2 x_i + N(0, 1)$

Проблема логит-пробит моделей

“Идеальное прогнозирование”:

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

ML оценки не существуют!

Объяснение с помощью ЧД

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, байесовский подход