

Семинар 26.

1. Винни-Пух знает, что мёд бывает правильный, $honey_i = 1$, и неправильный, $honey_i = 0$. Пчёлы также бывают правильные, $bee_i = 1$, и неправильные, $bee_i = 0$. По 100 своим попыткам добыть мёд Винни-Пух составил таблицу сопряженности:

	$honey_i = 1$	$honey_i = 0$
$bee_i = 1$	12	36
$bee_i = 0$	32	20

- Какие оценки коэффициентов получит Винни-Пух?
- Какой прогноз вероятности правильности мёда при встрече с неправильными пчёлами даёт логистическая модель? Как это число можно посчитать без расчета коэффициентов?
- Проверьте гипотезу о том, что правильность пчёл не оказывает влияние на правильность мёда с помощью тестов LR, LM и W.
- Держась в небе за воздушный шарик, Винни-Пух неожиданно понял, что перед ним неправильные пчёлы. Помогите ему оценить вероятность того, что они делают неправильный мёд.

Вспомним ММП!

Пусть требуется протестировать систему (нелинейных) ограничений относительно вектора неизвестных параметров

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ g_2(\theta) = 0 \\ \dots \\ g_r(\theta) = 0 \end{cases}$$

где $g_i(\theta)$ — функция, которая задаёт i -ое ограничение на вектор параметров θ , $i = 1, \dots, r$.

$$\frac{\partial g}{\partial \theta'} = \begin{pmatrix} \partial g_1 / \partial \theta' \\ \partial g_2 / \partial \theta' \\ \vdots \\ \partial g_r / \partial \theta' \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \dots & \frac{\partial g_1}{\partial \theta_k} \\ \frac{\partial g_2}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_2} & \dots & \frac{\partial g_2}{\partial \theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_r}{\partial \theta_1} & \frac{\partial g_r}{\partial \theta_2} & \dots & \frac{\partial g_r}{\partial \theta_k} \end{pmatrix}$$

$$\frac{\partial g'}{\partial \theta} = \begin{pmatrix} \frac{\partial g'_1}{\partial \theta} & \frac{\partial g'_2}{\partial \theta} & \dots & \frac{\partial g'_r}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_1} & \dots & \frac{\partial g_r}{\partial \theta_1} \\ \frac{\partial g_1}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_2} & \dots & \frac{\partial g_r}{\partial \theta_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_1}{\partial \theta_k} & \frac{\partial g_2}{\partial \theta_k} & \dots & \frac{\partial g_r}{\partial \theta_k} \end{pmatrix}$$

$$\frac{\partial \ell}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_k} \end{pmatrix}$$

$$I(\theta) = -E \left(\frac{\partial^2 \ell}{\partial \theta \partial \theta'} \right) = -E \begin{pmatrix} \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_k} \end{pmatrix} \text{ — информационная мат-}$$

рица Фишера

$\Theta_{UR} := \Theta$ — множество допустимых значений вектора неизвестных параметров без учёта ограничений

$\Theta_R := \{\theta \in \Theta : g(\theta) = 0\}$ — множество допустимых значений вектора неизвестных параметров с учётом ограничений

$\hat{\theta}_{UR} \in \Theta_{UR}$ — точка максимума функции ℓ на множестве Θ_{UR}

$\hat{\theta}_R \in \Theta_R$ — точка максимума функции ℓ на множестве Θ_R

Тогда для тестирования гипотезы H_0 можно воспользоваться одной из следующих ниже статистик.

$LR := -2(\ell(\hat{\theta}_R) - \ell(\hat{\theta}_{UR})) \stackrel{a}{\sim} \chi_r^2$ — статистика отношения правдоподобия

$W := g'(\hat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_{UR}) \cdot I^{-1}(\hat{\theta}_{UR}) \cdot \frac{\partial g}{\partial \theta}(\hat{\theta}_{UR}) \right]^{-1} g(\hat{\theta}_{UR}) \stackrel{a}{\sim} \chi_r^2$ — статистика Вальда

$LM := \left[\frac{\partial \ell}{\partial \theta}(\hat{\theta}_R) \right]' \cdot I^{-1}(\hat{\theta}_R) \cdot \left[\frac{\partial \ell}{\partial \theta}(\hat{\theta}_R) \right] \stackrel{a}{\sim} \chi_r^2$ — статистика множителей Лагранжа

2. Методом максимального правдоподобия оценили логит-модель

$$\hat{y}_i^* = 2 + 3x_i - 5z_i.$$

(а) Оцените вероятность того, что $y_i = 1$ для $\bar{x} = 15$, $\bar{z} = 3.5$.

(б) Оцените предельный эффект увеличения x на единицу на вероятность того, что $y_i = 1$ для $\bar{x} = 15$, $\bar{z} = 3.5$.

(с) При каком значении x предельный эффект увеличения x на единицу в точке $\bar{z} = 3.5$ будет максимальным?