Статистика

DS-поток

Лекция 14

8. Теория наилучших оценок

8.5 Оптимальные оценки

Оптимальные оценки

$X = (X_1, \ldots, X_n)$ — выборка из распр. $\mathsf{P} \in \mathscr{P} \in \{\mathsf{P}_\theta \mid \theta \in \Theta\}$;

$\mathcal{K} = \{$все несмещенные оценки параметра $\theta\}$

Задача: найти наилучшую в $\mathcal{K}$ в с/к-подходе.

Т.е. найти $\widehat{\theta} \in \mathcal{K}$, которая *для всех* $\theta \in \Theta$ дает минимум величины

$$MSE_{\widehat{\theta}}(\theta) = \mathsf{E}_\theta \left(\widehat{\theta} - \theta\right)^2 \underbrace{= \mathsf{D}_\theta \widehat{\theta}}_{\text{для оценок из } \mathcal{K}}$$

Такие оценки называются *оптимальными*.

Сравнение с эффективными оценками

Неравенство Рао-Крамера [E1-E4]

Пусть $\widehat{\theta}$ — несмещенная оценка $\tau(\theta)$. Тогда

$$\mathsf{D}_\theta \widehat{\theta} \geqslant \frac{(\tau'(\theta))^2}{I_X(\theta)}.$$

- Неравенство дает нижнюю границу дисперсию оптимальной оценки;
- Если на $\widehat{\theta}$ достигается равенство, то $\widehat{\theta}$ называется *эффективной*;
- Эффективные оценки являются оптимальными;
- Равенство может не достигаться, тогда эффективных оценок нет, а оптимальные могут существовать.

Достаточные статистики

Определение

Статистика $S(X)$ называется *достаточной* для семейства распределений $\{\mathsf{P}_\theta \mid \theta \in \Theta\}$, если условное распределение $\mathsf{P}_\theta(X \in B \mid S(X))$ не зависит от θ.

Смысл: вся информации о параметре θ, которая есть в выборке, содержится в достаточной статистике.

Следствие: в случае данных, поступающих последовательно, достаточно хранить только значения достаточных статистик.

Тривиальный пример достаточной статистики — вся выборка X.

Пример

Дана выборка $X_1, ..., X_n \sim Bern(\theta)$.

Какая информация есть в этой выборке?

1. $S(X) = \sum\limits_{i=1}^{n} X_i$ — количество успехов;
2. Порядок нулей и единиц.

   бесполезная информация, т.к. наблюдения независимы

Оценка $\widehat{\theta}_n = \overline{X}$ — функция от $S(X)$.

При поступлении нового объекта X_n оценка обновляется по правилу

$$\widehat{\theta}_n = \frac{(n-1)\widehat{\theta}_{n-1} + X_n}{n}.$$

Улучшение несмещенных оценок

Теорема (Колмогорова, Блекуэлла, Рао)

$\widehat{\theta}$ — несмещенная оценка $\tau(\theta)$, причем $\mathsf{E}_\theta \widehat{\theta}^2 < +\infty$;

$S(X)$ — достаточная статистика.

Тогда

1. $\theta^* = \mathsf{E}_\theta \left(\widehat{\theta} \,\middle|\, S(X) \right)$ тоже является несмещенной оценкой $\tau(\theta)$.
2. $\mathsf{D}_\theta \theta^* \leqslant \mathsf{D}_\theta \widehat{\theta} \quad \forall \theta \in \Theta$.

   Равенство возможно $\Leftrightarrow$ $\theta^* = \widehat{\theta}$ P_θ-п.н. $\forall \theta \in \Theta$,

   то есть $\widehat{\theta}$ изначально является $S(X)$-измеримой.

Если $\tau(\theta) \in \mathbb{R}^d$, $d > 1$, то $\mathsf{D}_\theta \theta^* \leqslant \mathsf{D}_\theta \widehat{\theta} \Leftrightarrow$ матрица $\mathsf{D}_\theta \widehat{\theta} - \mathsf{D}_\theta \theta^*$ неотр. опр..

Следствия:

1. $\theta^* = \mathsf{E}_\theta\left(\widehat{\theta}\,\middle|\,S(X)\right)$ не хуже исходной оценки $\widehat{\theta}$ в с/к подходе.
2. Если $\widehat{\theta}$ не является $S(X)$-измеримой,

   то θ^* лучше $\widehat{\theta}$ в с/к подходе.
3. Если θ^* — **единственная** несмещ. $S(X)$-измеримая оценка $\tau(\theta)$,

   то она и является оптимальной.

▶

Пусть $\xi(X)$ — несмещенная не $S(X)$-измеримая оценка $\tau(\theta)$.

Тогда $\xi^*(X) = \mathsf{E}_\theta\left(\xi(X) \mid S(X)\right)$:

- не хуже чем $\xi(X)$ в с/к подходе;
- несмещенная $S(X)$-измеримая оценка $\tau(\theta)$

  $\Rightarrow \quad \theta^* = \xi^*(X)$ в силу единственности.

Значит θ^* — оптимальная оценка $\tau(\theta)$. ■

Полные статистики

Единственность гарантирует свойство полноты.

Определение

Статистика $S(X)$ называется *полной* для семейства распределений $\{\mathsf{P}_\theta \mid \theta \in \Theta\}$, если выполнение свойство $\forall \theta \in \Theta : \mathsf{E}_\theta f(S(X)) = 0$ возможно только в случае $\forall \theta \in \Theta : f(S(X)) \overset{\mathsf{P}_\theta\text{-п.н.}}{=} 0$.

Смысл: несмещенной $S(X)$-измеримой оценкой нуля может быть только ноль.

Оптимальные оценки

Теорема (Об оптимальной оценке)

$S(X)$ — полная и достаточная статистика для $\{\mathsf{P}_\theta \mid \theta \in \Theta\}$;

Оценка $\theta^* = \varphi(S(X))$ — несмещенная $S(X)$-измеримая оценка $\tau(\theta)$.

Тогда θ^* — оптимальная оценка $\tau(\theta)$.

►

Согласно пред. следствию достаточно проверить, что

$\theta^* = \varphi(S(X))$ — единственная несмещенная $S(X)$-измеримая оценка $\tau(\theta)$.

Пусть $\psi(S(X))$ — тоже несмещенная оценка $\tau(\theta)$. Обозначим

$f(x) = \varphi(x) - \psi(x)$. Тогда $\mathsf{E}_\theta f(S(X)) = \varphi(S(X)) - \psi(S(X)) = 0 \ \forall \theta \in \Theta$.

Но $S(X)$ полная $\Longrightarrow$ P_θ-п.н. $\forall \theta \in \Theta: \ f(S(X)) = 0 = \varphi(S(X)) - \psi(S(X))$.

■

Следствия:

$S(X)$ — полная и достаточная статистика для $\{\mathrm{P}_\theta \mid \theta \in \Theta\}$.

Тогда

1. если θ^* — несмещенная оценка $\tau(\theta)$,

   то $\mathrm{E}_\theta\left(\theta^* | S(X)\right)$ — оптимальная оценка $\tau(\theta)$;

2. если θ_1^*, θ_2^* — оптимальные оценки $\tau_1(\theta), \tau_2(\theta)$,

   то $a\theta_1^* + b\theta_2^*$ — оптимальная оценка $a\tau_1(\theta) + b\tau_2(\theta)$;

3. если $\tau(\theta) = (\tau_1(\theta), ..., \tau_k(\theta)) \in \mathbb{R}^k$ и θ_j^* — оптим. оценка $\tau_j(\theta)$,

   то $(\theta_1^*, ..., \theta_k^*)$ — оптимальная оценка вектора $\tau(\theta)$.

Алгоритм поиска оптимальных оценок

1. Найти $S(X)$ — полную и достаточную статистику в данной модели;
2. Решить уравнение несмещенности $\mathsf{E}_\theta \varphi(S(X)) = \tau(\theta)$ относительно φ. Оценка $\theta^* = \varphi(S(X))$ будет оптимальной оценкой $\tau(\theta)$ согласно теореме об оптимальной оценке.

Экспоненциальное семейство

Пусть $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$, причем плотность $p_\theta(x) = \frac{g(x)}{h(\theta)} e^{a(\theta)^T u(x)}$.

Теорема

Если множество Θ телесно (т.е. содержит внутренние точки), а функция $a(\theta)$ непрерывна и содержит линейно независимые компоненты, то статистика $S(X) = \sum_{i=1}^{n} u(X_i)$ является полной и достаточной для семейства $\mathscr{P}$.

Доказательства теорем

Теорема (Колмогорова, Блекуэлла, Рао)

$\widehat{\theta}$ — несмещенная оценка $\tau(\theta)$;

$S(X)$ — достаточная статистика, причем $\mathsf{E}_\theta \widehat{\theta}^2 < +\infty$.

Тогда

1. $\theta^* = \mathsf{E}_\theta \left(\widehat{\theta} \,\middle|\, S(X) \right)$ тоже является несмещенной оценкой $\tau(\theta)$.
2. $\mathsf{D}_\theta \theta^* \leqslant \mathsf{D}_\theta \widehat{\theta} \quad \forall \theta \in \Theta$.

   Равенство возможно $\Leftrightarrow$ $\theta^* = \widehat{\theta}$ P_θ-п.н. $\forall \theta \in \Theta$,

   то есть $\widehat{\theta}$ изначально является $S(X)$-измеримой.

Доказываемое утверждение:

$\widehat{\theta}$ — несмещенная оценка $\tau(\theta)$;

$S(X)$ — достаточная статистика, причем $\mathsf{E}_\theta \widehat{\theta}^2 < +\infty$.

Тогда $\theta^* = \mathsf{E}_\theta \left(\widehat{\theta} \,\middle|\, S(X) \right)$ тоже является несмещенной оценкой $\tau(\theta)$.

►

$S(X)$ — достаточная $\Rightarrow \mathsf{P}_\theta(X \in B \mid S(X))$ не зависит от θ.

$\Rightarrow \mathsf{E}_\theta \left(\widehat{\theta} \,\middle|\, S(X) \right)$ тоже не зависит от θ (м.о. по условному распр.)

$\Rightarrow \theta^*$ — действительно оценка.

$$\mathsf{E}_\theta \theta^* = \mathsf{E}_\theta \left[\mathsf{E}_\theta \left(\widehat{\theta} \,\middle|\, S(X) \right) \right] = \mathsf{E}_\theta \widehat{\theta} = \tau(\theta)$$

$\Rightarrow \theta^*$ — несмещенная оценка $\tau(\theta)$.

■

Доказываемое утверждение:

$\widehat{\theta}$ — несмещенная оценка $\tau(\theta)$;

$S(X)$ — достаточная статистика, причем $\mathsf{E}_\theta \widehat{\theta}^2 < +\infty$.

Тогда $\mathsf{D}_\theta \theta^* \leqslant \mathsf{D}_\theta \widehat{\theta} \quad \forall \theta \in \Theta$.

- (для $\tau(\theta) \in \mathbb{R}$)

$$\mathsf{D}_\theta \widehat{\theta} = \mathsf{E}_\theta \left(\widehat{\theta} - \tau(\theta)\right)^2 = \mathsf{E}_\theta \left(\widehat{\theta} - \theta^* + \theta^* - \tau(\theta)\right)^2 =$$

$$\underbrace{\mathsf{E}_\theta \left(\widehat{\theta} - \theta^*\right)^2}_{\geqslant 0} + \mathsf{D}_\theta \theta^* + \underbrace{2\mathsf{E}_\theta \left(\widehat{\theta} - \theta^*\right)(\theta^* - \tau(\theta))}_{=0} \geqslant \mathsf{D}_\theta \theta^*.$$

$$\mathsf{E}_\theta \left(\widehat{\theta} - \theta^*\right)\{\theta^* - \tau(\theta)\} = \mathsf{E}_\theta \left(\mathsf{E}_\theta \left[\left(\widehat{\theta} - \theta^*\right)\{\theta^* - \tau(\theta)\} \middle| S(X)\right]\right) =$$

$$= \mathsf{E}_\theta \left(\{\theta^* - \tau(\theta)\}\, \mathsf{E}_\theta \left[\widehat{\theta} - \theta^* \middle| S(X)\right]\right) = \mathsf{E}_\theta \left((\theta^* - \tau(\theta)) \cdot 0\right) = 0$$

Воспользовались $S(X)$-измеримостью θ^* и свойствами УМО.

■

Доказываемое утверждение:

$\widehat{\theta}$ — несмещенная оценка $\tau(\theta)$;

$S(X)$ — достаточная статистика, причем $\mathsf{E}_\theta \widehat{\theta}^2 < +\infty$.

Тогда $\mathsf{D}_\theta \theta^* = \mathsf{D}_\theta \widehat{\theta} \quad \forall \theta \in \Theta \quad \Leftrightarrow \quad \theta^* = \widehat{\theta}$ P_θ-п.н. $\forall \theta \in \Theta$,

то есть $\widehat{\theta}$ изначально является $S(X)$-измеримой.

- (для $\tau(\theta) \in \mathbb{R}$)

Из док-ва предыдущего утверждения равенство возможно $\Leftrightarrow$

$\mathsf{E}_\theta \left(\widehat{\theta} - \theta^*\right)^2 = 0 \quad \forall \theta \in \Theta \quad \Leftrightarrow$

$\widehat{\theta} = \theta^*$ P_θ-п.н. $\forall \theta \in \Theta \quad \Leftrightarrow$

$\widehat{\theta} = \mathsf{E}_\theta \left(\widehat{\theta} \,\middle|\, S(X)\right)$ P_θ-п.н. $\forall \theta \in \Theta \quad \Leftrightarrow$

$\widehat{\theta}$ является $S(X)$-измеримой.

$\blacksquare$

Оптимальные оценки в гауссовской линейной модели

Гауссовская линейная модель

$$Y = X\theta + \varepsilon,$$

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \dots & & \\ x_{n1} & \dots & x_{nd} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \dots \\ \theta_d \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Предполагаем, что $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.

$\widehat{\theta} = \left(X^T X\right)^{-1} X^T Y$ — несмещенная оценка θ

$\widehat{\sigma}^2 = \frac{1}{n-d} \left\| Y - X\widehat{\theta} \right\|^2$ — несмещенная оценка σ^2

Достаточная статистика

$L(X) = \{X\theta \mid \theta \in \mathbb{R}^d\}$ — подпространство в $\mathbb{R}^n$, порождаемое X.

Утверждение

$S(Y) = \left(\mathrm{proj}_{L(X)} Y, \left\|\mathrm{proj}_{L^\perp(X)} Y\right\|^2\right)$ — *достаточная статистика.*

▶

Запишем плотность $Y \sim \mathcal{N}(X\theta, \sigma^2 I_n)$ // $c = (2\pi\sigma^2)^{-n/2}$

$$p(y) = c \cdot \exp\left(-\frac{1}{2\sigma^2}\sum\left(Y_i - x_i^T\theta\right)^2\right) = c \cdot \exp\left(-\frac{1}{2\sigma^2}\|Y - X\theta\|^2\right) =$$

$$= c \cdot \exp\left(-\frac{1}{2\sigma^2}\left[\left\|\mathrm{proj}_{L(X)}(Y - X\theta)\right\|^2 + \left\|\mathrm{proj}_{L^\perp(X)}(Y - X\theta)\right\|^2\right]\right) =$$

$$= c \cdot \exp\left(-\frac{1}{2\sigma^2}\left[\left\|\mathrm{proj}_{L(X)}Y - X\theta\right\|^2 + \left\|\mathrm{proj}_{L^\perp(X)}Y\right\|^2\right]\right)$$

По критерию факторизации $S(Y)$ — достаточная. ■

Оптимальные оценки

Утверждение

1. *$S(Y)$ — полная статистика (б/д);*
2. *$\widehat{\theta} = \left(X^T X\right)^{-1} X^T Y$ — оптимальная оценка θ;*
3. *$\widehat{\sigma}^2 = \frac{1}{n-d} \left\| Y - X\widehat{\theta} \right\|^2$ — оптимальная оценка σ^2.*

- Док-во: Обе несмещенные и явл. функциями от $S(Y)$. ■

Утверждение

Если не предполагать нормальность ошибки,

то $\widehat{\theta}$ — наилучшая оценка в с/к-подходе в классе всех несмещенных оценок, которые являются линейными по Y.

8. Теория наилучших оценок

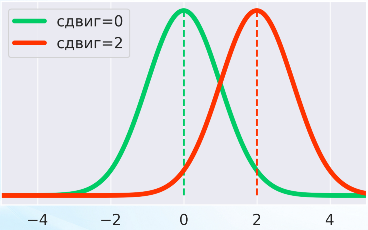
8.6 Оценки параметров масштаба и сдвига

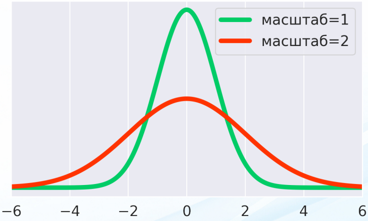
Сдвиг и масштаб

$\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ — семейство непрер. распр. с плотностью $p_\theta(x)$

θ — параметр сдвига, если $p_\theta(x) = p_0(x-\theta)$.

θ — параметр масштаба, если $p_\theta(x) = \frac{1}{\theta} p_1(x/\theta)$ и $\Theta \subset (0, +\infty)$





Примеры:

$\mathscr{P} = \{\mathcal{N}(a, \sigma^2) \mid a \in \mathbb{R}, \sigma > 0\}$:

a — параметр сдвига

σ — параметр масштаба.

$\mathscr{P} = \{U[0, \theta] \mid \theta > 0\}$:

θ — параметр масштаба.

Параметр сдвига

Естественно предполагать $\Theta = \mathscr{X} = \mathbb{R}^d$ и распр. P_0 известно.

$X = (X_1, ..., X_n)$ — выборка из распределения P_θ.

$\Rightarrow\ X + c = (X_1 + c, ..., X_n + c)$ имеют распределение $\mathsf{P}_{\theta - c}$.

$\Rightarrow$ рассматриваем *эквивариантные* оценки относительно сдвига:
$\mathcal{K} = \left\{ \widehat{\theta} \;\middle|\; \forall c \in \mathscr{X} : \widehat{\theta}(X + c) = \widehat{\theta}(X) + c \right\}$

Теорема

Пусть P_0 *непрерывна и имеет плотность* $p_0(x)$*. Тогда оценка Питмена*

$$\widehat{\theta} = \frac{\int\limits_{\mathbb{R}^d} t\, L_X(t)\, dt}{\int\limits_{\mathbb{R}^d} L_X(t)\, dt} = \frac{\int\limits_{\mathbb{R}^d} t\, p_0(X - t)\, dt}{\int\limits_{\mathbb{R}^d} p_0(X - t)\, dt}$$

является несмещенной оценкой θ *и единственной наилучшей оценкой в среднеквадратичном подходе в классе всех эквивариантных относительно сдвига оценок* θ*.*

Параметр сдвига (пример)

Найдем оценку Питмена в модели $U[\theta, \theta+1]$.

$$L_X(\theta) = \prod_{i=1}^{n} I\{\theta \leqslant X_i \leqslant \theta+1\} = I\{X_{(1)} \geqslant \theta, X_{(n)} \leqslant \theta+1\} =$$
$$= I\{X_{(n)} - 1 \leqslant \theta \leqslant X_{(1)}\}$$

$$\widehat{\theta} = \frac{\int\limits_{\mathbb{R}} t\, L_X(t)\, dt}{\int\limits_{\mathbb{R}} L_X(t)\, dt} = \frac{\int\limits_{\mathbb{R}} t\, I\{X_{(n)} - 1 \leqslant t \leqslant X_{(1)}\}\, dt}{\int\limits_{\mathbb{R}} I\{X_{(n)} - 1 \leqslant t \leqslant X_{(1)}\}\, dt} = \frac{\int\limits_{X_{(n)}-1}^{X_{(1)}} t\, dt}{\int\limits_{X_{(n)}-1}^{X_{(1)}} dt} =$$
$$= \frac{t}{2}\Big|_{X_{(n)}-1}^{X_{(1)}} = \frac{X_{(1)} + X_{(n)} - 1}{2}.$$

Замечание. ОМП — любая статистика из отрезка $\left[X_{(n)} - 1, X_{(1)}\right]$.

Оценка Питмена есть середина этого отрезка.

Параметр масштаба

Естественно предполагать $\sigma \in \Sigma = (0, +\infty)$ и распр. P_1 известно.

$X = (X_1, ..., X_n)$ — выборка из распределения P_σ.

$\Rightarrow$ $cX = (cX_1, ..., cX_n)$ имеют распределение $\mathsf{P}_{\sigma/c}$.

$\Rightarrow$ рассматриваем *эквивариантные* оценки относительно масштаба:

$\mathcal{K} = \{\widehat{\sigma} \mid \forall c \in \mathscr{X} : \widehat{\sigma}(cX) = c\widehat{\sigma}(X)\}$

Теорема

Пусть P_1 *непрерывна и имеет плотность* $p_1(x)$. *Тогда оценка*

$$\widehat{\sigma} = \frac{\int\limits_0^{+\infty} s^{-n-k-1} p_1(X/s) ds}{\int\limits_0^{+\infty} s^{-n-2k-1} p_1(X/s) ds} = \frac{\int\limits_0^{+\infty} u^{n+k-1} p_1(uX) du}{\int\limits_0^{+\infty} u^{n+2k-1} p_1(uX) du}$$

является единственной наилучшей оценкой σ^k *в классе всех эквивариантных относительно масштаба оценок* σ^k *в равномерном подходе с функцией риска* $\mathsf{E}_\sigma \left(\frac{\widehat{\sigma}}{\sigma} - 1\right)^2$.


ВСЁ!