

Статистика ФИВТ ПМИ

Прикладной поток

Лекция 15

8. Теория наилучших оценок

8.4 Оптимальные оценки

Оптимальные оценки

$X = (X_1, \dots, X_n)$ — выборка из распр. $P \in \mathcal{P} \in \{P_\theta \mid \theta \in \Theta\}$;

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

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

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

$$MSE_{\hat{\theta}}(\theta) = E_\theta (\hat{\theta} - \theta)^2 = \underbrace{D_\theta \hat{\theta}}_{\text{для оценок из } \mathcal{K}}$$

Такие оценки называются *оптимальными*.

Сравнение с эффективными оценками

Неравенство Рао-Крамера [E1-E4]

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

$$D_{\theta}\hat{\theta} \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}.$$

- ▶ Неравенство дает нижнюю границу дисперсию оптимальной оценки;
- ▶ Если на $\hat{\theta}$ достигается равенство, то $\hat{\theta}$ 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 $\{P_\theta \mid \theta \in \Theta\}$, если условное распределение $P_\theta(X \in B \mid S(X))$ 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 $X_1, \dots, X_n \sim \text{Bern}(\theta)$.

Какая информация есть в этой выборке?

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

бесполезная информация, т.к. наблюдения независимы

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

При поступлении нового объекта X_n оценка обновляется по правилу

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

Улучшение несмещенных оценок

Теорема (Колмогорова, Блекуэлла, Рао)

$\hat{\theta}$ — несмещенная оценка $\tau(\theta)$, причем $E_{\theta}\hat{\theta}^2 < +\infty$;

$S(X)$ — достаточная статистика.

Тогда

1. $\theta^* = E_{\theta}(\hat{\theta} \mid S(X))$ тоже является несмещенной оценкой $\tau(\theta)$.
2. $D_{\theta}\theta^* \leq D_{\theta}\hat{\theta} \quad \forall \theta \in \Theta$.

Равенство возможно $\Leftrightarrow \theta^* = \hat{\theta}$ P_{θ} -п.н. $\forall \theta \in \Theta$,

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

Если $\tau(\theta) \in \mathbb{R}^d, d > 1$, то $D_{\theta}\theta^* \leq D_{\theta}\hat{\theta} \Leftrightarrow$ матрица $D_{\theta}\hat{\theta} - D_{\theta}\theta^*$ неотр. опр..

Следствия:

1. $\theta^* = E_{\theta}(\hat{\theta} \mid S(X))$ не хуже исходной оценки $\hat{\theta}$ в с/к подходе.
2. Если $\hat{\theta}$ не является $S(X)$ -измеримой, то θ^* лучше $\hat{\theta}$ 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 $\xi^*(X) = E_{\theta}(\xi(X) \mid S(X))$:

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

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

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



Полные статистики

Единственность гарантирует свойство полноты.

Определение

Статистика $S(X)$ называется *полной* для семейства распределений $\{P_\theta \mid \theta \in \Theta\}$, если выполнение свойства $\forall \theta \in \Theta : E_\theta f(S(X)) = 0$ возможно только в случае $\forall \theta \in \Theta : f(S(X)) \stackrel{P_\theta\text{-п.н.}}{=} 0$.

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

Пример. Пусть $X_1, \dots, X_n \sim U[0, \theta]$. Найти полную статистику.

Оптимальные оценки

Теорема (Об оптимальной оценке)

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

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 $E_\theta f(S(X)) = \varphi(S(X)) - \psi(S(X)) = 0 \quad \forall \theta \in \Theta$.

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

Следствия:

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

Тогда

1. если θ^* — несмещенная оценка $\tau(\theta)$,
то $E_\theta(\theta^* | S(X))$ — 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 $\tau(\theta) = (\tau_1(\theta), \dots, \tau_k(\theta)) \in \mathbb{R}^k$ и θ_j^* — оптим. оценка $\tau_j(\theta)$,
то $(\theta_1^*, \dots, \theta_k^*)$ — оптимальная оценка вектора $\tau(\theta)$.

Алгоритм поиска оптимальных оценок

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

Пример. Пусть $X_1, \dots, X_n \sim U[0, \theta]$. Найти оптимальную оценку.



Экспоненциальное семейство

Пусть $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, 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 $\mathcal{P}$.



Доказательства теорем

Теорема (Колмогорова, Блекуэлла, Рао)

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

$S(X)$ — достаточная статистика, причем $E_{\theta}\hat{\theta}^2 < +\infty$.

Тогда

1. $\theta^* = E_{\theta}(\hat{\theta} \mid S(X))$ тоже является несмещенной оценкой $\tau(\theta)$.
2. $D_{\theta}\theta^* \leq D_{\theta}\hat{\theta} \quad \forall \theta \in \Theta$.

Равенство возможно $\Leftrightarrow \theta^* = \hat{\theta}$ P_{θ} -п.н. $\forall \theta \in \Theta$,
то есть $\hat{\theta}$ изначально является $S(X)$ -измеримой.

Доказываемое утверждение:

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

$S(X)$ — достаточная статистика, причем $E_{\theta}\hat{\theta}^2 < +\infty$.

Тогда $\theta^* = E_{\theta} \left(\hat{\theta} \mid S(X) \right)$ тоже является несмещенной оценкой $\tau(\theta)$.



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

$\Rightarrow E_{\theta} \left(\hat{\theta} \mid S(X) \right)$ тоже не зависит от θ (м.о. по условному распр.)

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

$$E_{\theta}\theta^* = E_{\theta} \left[E_{\theta} \left(\hat{\theta} \mid S(X) \right) \right] = E_{\theta}\hat{\theta} = \tau(\theta)$$

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





Доказываемое утверждение:

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

$S(X)$ — достаточная статистика, причем $E_{\theta}\hat{\theta}^2 < +\infty$.

Тогда $D_{\theta}\theta^* \leq D_{\theta}\hat{\theta} \quad \forall \theta \in \Theta$.

► (для $\tau(\theta) \in \mathbb{R}$)

$$\begin{aligned} D_{\theta}\hat{\theta} &= E_{\theta} \left(\hat{\theta} - \tau(\theta) \right)^2 = E_{\theta} \left(\hat{\theta} - \theta^* + \theta^* - \tau(\theta) \right)^2 = \\ &= \underbrace{E_{\theta} \left(\hat{\theta} - \theta^* \right)^2}_{\geq 0} + D_{\theta}\theta^* + \underbrace{2E_{\theta} \left(\hat{\theta} - \theta^* \right) (\theta^* - \tau(\theta))}_{=0} \geq D_{\theta}\theta^*. \end{aligned}$$

$$\begin{aligned} E_{\theta} \left(\hat{\theta} - \theta^* \right) \{ \theta^* - \tau(\theta) \} &= E_{\theta} \left(E_{\theta} \left[\left(\hat{\theta} - \theta^* \right) \{ \theta^* - \tau(\theta) \} \mid S(X) \right] \right) = \\ &= E_{\theta} \left(\{ \theta^* - \tau(\theta) \} E_{\theta} \left[\hat{\theta} - \theta^* \mid S(X) \right] \right) = E_{\theta} \left((\theta^* - \tau(\theta)) \cdot 0 \right) = 0 \end{aligned}$$

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

Доказываемое утверждение:

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

$S(X)$ — достаточная статистика, причем $E_{\theta}\hat{\theta}^2 < +\infty$.

Тогда $D_{\theta}\theta^* = D_{\theta}\hat{\theta} \quad \forall \theta \in \Theta \Leftrightarrow \theta^* = \hat{\theta} \text{ P}_{\theta}\text{-п.н. } \forall \theta \in \Theta$,
то есть $\hat{\theta}$ изначально является $S(X)$ -измеримой.

► (для $\tau(\theta) \in \mathbb{R}$)

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

$$E_{\theta} \left(\hat{\theta} - \theta^* \right)^2 = 0 \quad \forall \theta \in \Theta \Leftrightarrow$$

$$\hat{\theta} = \theta^* \text{ P}_{\theta}\text{-п.н. } \forall \theta \in \Theta \Leftrightarrow$$

$$\hat{\theta} = E_{\theta} \left(\hat{\theta} \mid S(X) \right) \text{ P}_{\theta}\text{-п.н. } \forall \theta \in \Theta \Leftrightarrow$$

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



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

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

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

Достаточная статистика

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

Утверждение

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



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

$$\begin{aligned} p(y) &= c \cdot \exp \left(-\frac{1}{2\sigma^2} \sum (Y_i - x_i^T \theta)^2 \right) = c \cdot \exp \left(-\frac{1}{2\sigma^2} \|Y - X\theta\|^2 \right) = \\ &= c \cdot \exp \left(-\frac{1}{2\sigma^2} \left[\left\| \text{proj}_{L(X)} (Y - X\theta) \right\|^2 + \left\| \text{proj}_{L^\perp(X)} (Y - X\theta) \right\|^2 \right] \right) = \\ &= c \cdot \exp \left(-\frac{1}{2\sigma^2} \left[\left\| \text{proj}_{L(X)} Y - X\theta \right\|^2 + \left\| \text{proj}_{L^\perp(X)} Y \right\|^2 \right] \right) \end{aligned}$$

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

Оптимальные оценки

Утверждение

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

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

Утверждение

Если не предполагать нормальность ошибки,
то $\hat{\theta}$ — наилучшая оценка в с/к-подходе в классе всех несмещенных
оценок, которые являются линейными по Y .

8. Теория наилучших оценок

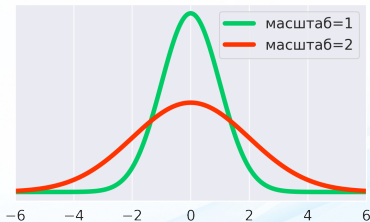
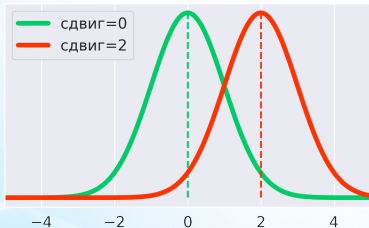
8.5 Оценки параметров масштаба и сдвига

Сдвиг и масштаб

$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ — 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:

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

a — параметр сдвига

σ — параметр масштаба.

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

θ — параметр масштаба.

Параметр сдвига

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

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

$\Rightarrow X + c = (X_1 + c, \dots, X_n + c)$ имеют распределение $P_{\theta+c}$.

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

$$\mathcal{K} = \left\{ \hat{\theta} \mid \forall c \in \mathcal{X} : \hat{\theta}(X + c) = \hat{\theta}(X) + c \right\}$$

Теорема

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

$$\hat{\theta} = \frac{\int_{\mathbb{R}^d} t L_X(t) dt}{\int_{\mathbb{R}^d} L_X(t) dt} = \frac{\int_{\mathbb{R}^d} t p_0(X - t) dt}{\int_{\mathbb{R}^d} p_0(X - t) dt}$$

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 $U[\theta, \theta + 1]$.

$$\begin{aligned} L_X(\theta) &= \prod_{i=1}^n I\{\theta \leq X_i \leq \theta + 1\} = I\{X_{(1)} \geq \theta, X_{(n)} \leq \theta + 1\} = \\ &= I\{X_{(n)} - 1 \leq \theta \leq X_{(1)}\} \end{aligned}$$

$$\begin{aligned} \hat{\theta} &= \frac{\int_{\mathbb{R}} t L_X(t) dt}{\int_{\mathbb{R}} L_X(t) dt} = \frac{\int_{\mathbb{R}} t I\{X_{(n)} - 1 \leq t \leq X_{(1)}\} dt}{\int_{\mathbb{R}} I\{X_{(n)} - 1 \leq t \leq X_{(1)}\} dt} = \\ &= \frac{\int_{X_{(n)}-1}^{X_{(1)}} t dt}{\int_{X_{(n)}-1}^{X_{(1)}} dt} = \frac{X_{(1)} + X_{(n)} - 1}{2}. \end{aligned}$$

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

Оценка Питмена есть середина этого отрезка.

Параметр масштаба

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

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

$\Rightarrow cX = (cX_1, \dots, cX_n)$ имеют распределение $P_{\sigma/c}$.

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

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

Теорема

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

$$\hat{\sigma} = \frac{\int_0^{+\infty} s^{-n-k-1} p_1(X/s) ds}{\int_0^{+\infty} s^{-n-2k-1} p_1(X/s) ds} = \frac{\int_0^{+\infty} u^{n+k-1} p_1(uX) du}{\int_0^{+\infty} u^{n+2k-1} p_1(uX) du}$$

является единственной наилучшей оценкой σ^k в классе всех equivариантных относительно масштаба оценок σ^k в равномерном подходе с функцией риска $E_\sigma \left(\frac{\hat{\sigma}}{\sigma} - 1 \right)^2$.



BCÈ !