МИНИМАКСНОЕ ОЦЕНИВАНИЕ В НЕОПРЕДЕЛЕННО-СТОХАСТИЧЕСКИХ МОДЕЛЯХ ЛИНЕЙНОЙ РЕГРЕССИИ

Семенихин Константин Владимирович

Кафедра теории вероятностей
Факультет прикладной математики и физики
Московский авиационный институт

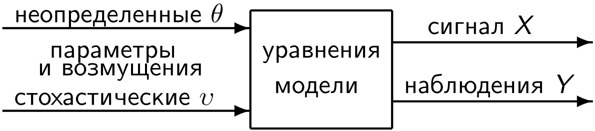
Пятая традиционная всероссийская молодежная летняя школа
«Управление, информация и оптимизация»
Солнечногорск, 17 июня 2013 г.

План лекции

1. Постановка задачи оценивания
   - Неопределенно-стохастическая модель линейной регрессии
   - Минимаксный подход
2. Проблема оптимальности линейных оценок
   - Равномерно оптимальные оценки
   - Минимаксные оценки в гауссовском случае
   - Минимаксные оценки при неизвестном распределении
   - Случай ограниченных параметров
3. Синтез минимаксных линейных оценок
   - Метод двойственной оптимизации
   - Метод полуопределенного программирования

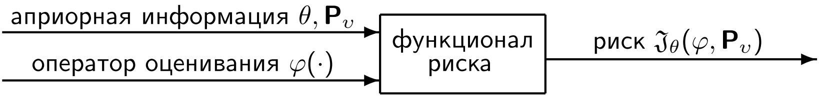
Постановка задачи оценивания

- **Модель наблюдения** — уравнения зависимости полезного сигнала X и наблюдений Y от параметров и возмущений + априорная информация: $\theta \in \Theta$, $\mathbf{P}_v \in \mathcal{P}_v$



- **Допустимые операторы оценивания** φ — доступные средства извлечения информации из данных эксперимента

- **Критерий оценивания** — правило выбора оператора оценивания на основе имеющейся априорной информации



Неопределенные системы

$X = a\theta, \quad Y = A\theta + \varepsilon$

неопределенные параметры $\theta \in \Theta$

неопределенные ошибки $\varepsilon \in \mathrm{E}$

Статистические модели

$X = a\theta, \quad Y = A\theta + \varepsilon$

неопределенные параметры $\theta \in \Theta$

случайные ошибки $\varepsilon \sim \mathbf{P}_\varepsilon$

Неопределенно-стохастические модели

$X = a\theta + b\upsilon, \quad Y = A\theta + B\upsilon$

θ, υ — векторы параметров, возмущений, ошибок и т.д.

θ — неопределенный вектор: $\theta \in \Theta$

υ — случайный вектор: $\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon$

Стохастические модели

$X = a\xi, \quad Y = A\xi + \varepsilon$

случайные параметры $\xi \sim \mathbf{P}_\xi$

случайные ошибки $\varepsilon \sim \mathbf{P}_\varepsilon$

Пусть $X = a\theta + bv \in \mathbb{R}^m, \quad Y = A\theta + Bv \in \mathbb{R}^n$.

Классы операторов оценивания:

- $\mathcal{B}$ содержит все борелевские отображения $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$;
- $\mathcal{B}_0$ состоит из несмещенных операторов, т.е.

$$\mathbf{E}_\theta\{\varphi(Y) - X\} = 0 \quad \forall\,\theta;$$

- $\mathcal{L}$ содержит все линейные операторы $L\colon \mathbb{R}^n \to \mathbb{R}^m$;
- $\mathcal{L}_0 = \mathcal{L} \cap \mathcal{B}_0 = \{L \in \mathcal{L}\colon\, LA = a\}$ — класс линейных несмещенных операторов оценивания, выполнено условие идентифицируемости

$$\mathcal{L}_0 \neq \varnothing \quad \iff \quad \ker[A] \subseteq \ker[a] \quad \iff \quad aA^+A = a.$$

Функционал риска:

- $\mathfrak{J}_\theta(\varphi, \mathbf{P}_v) = \mathbf{E}_\theta|\varphi(Y) - X|^2$ — с.к. риск ($|\cdot|$ — евклидова норма).

Несколько оптимизационных постановок

Пусть $\mathcal{F}$ — класс допустимых операторов оценивания.

Определение 1 (Оптимальный подход)

$\hat{X} = \hat{\varphi}(Y)$ — равномерно оптимальная оценка, если $\forall\, \varphi \in \mathcal{F}$

$$\mathfrak{J}_\theta(\hat{\varphi}, \mathbf{P}_\upsilon) \leq \mathfrak{J}_\theta(\varphi, \mathbf{P}_\upsilon) \quad \forall\, \theta \in \Theta \quad \forall\, \mathbf{P}_\upsilon \in \mathcal{P}_\upsilon.$$

Определение 2 (Минимаксный подход)

$\hat{X} = \hat{\varphi}(Y)$ — минимаксная оценка, если для любого $\varphi \in \mathcal{F}$

$$\sup_{\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon} \sup_{\theta \in \Theta} \mathfrak{J}_\theta(\hat{\varphi}, \mathbf{P}_\upsilon) \leq \sup_{\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon} \sup_{\theta \in \Theta} \mathfrak{J}_\theta(\varphi, \mathbf{P}_\upsilon).$$

Определение 3 (Минимаксно-асимптотический подход)

$\hat{X}^{(n)} = \hat{\varphi}^{(n)}(Y^{(n)})$ — минимаксно-асимптотическая оценка, если для любого $\varphi = \{\varphi^{(n)}\} \in \mathcal{F}$

$$\sup_{\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon} \lim_{n \to \infty} n\, \mathfrak{J}_\theta(\hat{\varphi}^{(n)}, \mathbf{P}_\upsilon) \leq \sup_{\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon} \lim_{n \to \infty} n\, \mathfrak{J}_\theta(\varphi^{(n)}, \mathbf{P}_\upsilon) \quad \forall\, \theta \in \Theta.$$

Пример 1 (Гауссовская регрессия)

Если совместное распределение оцениваемого вектора X и вектора наблюдений Y — гауссовское $\mathcal{N}\left(0, \begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix}\right)$, то линейная оценка $\boxed{\hat{X} = R_{XY} R_Y^{+} Y}$ оптимальна на классе всех измеримых оценок $\widetilde{X} = \varphi(Y)$, $\varphi \in \mathcal{B}$, так как $\hat{X} = \mathbf{E}\{X \mid Y\}$ по теореме о нормальной корреляции.

Пример 2 (Линейная регрессия при нормальных ошибках)

В модели $X = a\theta$, $Y = A\theta + \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, R_\varepsilon)$, $R_\varepsilon \succ O$, оценка Гаусса—Маркова $\boxed{\hat{X} = a\left(A^* R_\varepsilon^{-1} A\right)^{+} A^* R_\varepsilon^{-1} Y}$ равномерно оптимальна на классе всех несмещенных оценок $\widetilde{X} = \varphi(Y)$, $\varphi \in \mathcal{B}_0$, в силу неравенства Рао—Крамера.

Вопрос

Как распространить результаты примеров 1 и 2 на случай обобщенной модели

$$\begin{cases} X = a\theta + \xi, \\ Y = A\theta + \eta, \end{cases} \qquad \theta \in \mathbb{R}^p, \quad \begin{bmatrix} \xi \\ \eta \end{bmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix}\right)?$$

Лемма 1 (Н.л.н. оценка в обобщенной модели)

Наилучшая линейная несмещенная (н.л.н.) оценка $\hat{X}$ вектора X по вектору Y существует и имеет вид

- $\hat{X} = a\hat{\theta} + \hat{\xi}$,
- $\hat{\theta} = A^+[I - R_Y(QR_YQ)^+]Y$ — *н.л.н. оценка вектора неопределенных параметров* θ, $Q = I - AA^+$ — *ортопроектор на* $\mathrm{im}[A]^\perp$,
- $\hat{\xi} = R_{XY}(QR_YQ)^+Y = \mathbf{E}\{\xi \mid QY\}$, *где* $QY = Y - AA^+Y$ — *остаточный вектор.*

Равномерная оптимальность: утверждение

Теорема 1

В модели наблюдения

$$\begin{cases} X = a\theta + \xi, \\ Y = A\theta + \eta, \end{cases} \qquad \theta \in \mathbb{R}^p, \quad \begin{bmatrix} \xi \\ \eta \end{bmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix}\right)$$

наилучшая линейная несмещенная оценка $\hat{X}$ является равномерно оптимальной на классе всех несмещенных оценок.

Лемма 2 (Критерий оптимальности)

Несмещенная оценка $\hat{X}$ равномерно оптимальна тогда и только тогда, когда для всякой несмещенной оценки $\widetilde{X}$ выполнено

$$\mathbf{E}_\theta \langle \hat{X} - X, \hat{X} - \widetilde{X} \rangle = 0 \quad \forall \theta \in \mathbb{R}^p.$$

Лемма 3 (Основная)

Ошибка наилучшей линейной несмещенной оценки $\Delta = \hat{X} - X$ *и вектор наблюдений* Y *условно независимы относительно н.л.н. оценки* $\hat{\theta}$ *вектора неопределенных параметров* θ.

<u>Доказательство леммы.</u> С помощью теоремы о нормальной корреляции вычисляем условную ковариацию $\mathbf{cov}\{\Delta, Y \mid \hat{\theta}\} =$
$= \mathbf{cov}\{\Delta, Y\} - \mathbf{cov}\{\Delta, \hat{\theta}\}\left(\mathbf{cov}\{\hat{\theta}, \hat{\theta}\}\right)^{+}\mathbf{cov}\{\hat{\theta}, Y\} = O.$

Доказательство.

$$\mathbf{E}_\theta \langle \Delta, \hat{X} - \widetilde{X} \rangle = \mathbf{E}_\theta \langle \mathbf{E}_\theta\{\Delta \mid \hat{\theta}\}, \mathbf{E}_\theta\{\hat{X} - \widetilde{X} \mid \hat{\theta}\}\rangle = \dots$$

- Существует линейный оператор Λ: $\mathbf{E}_\theta\{\Delta \mid \hat{\theta}\} = \Lambda(\hat{\theta} - \theta)$.
- Найдется несмещенная оценка $\widetilde{\theta}$: $\Lambda^*\mathbf{E}_\theta\{\hat{X} - \widetilde{X} \mid \hat{\theta}\} = \hat{\theta} - \widetilde{\theta}$.

$$\dots = \mathbf{E}_\theta \langle \hat{\theta} - \theta, \hat{\theta} - \widetilde{\theta} \rangle = 0. \qquad \square$$

Пример 3 (Теорема Ходжеса—Лемана)

В модели $Y = A\theta + \varepsilon$, где θ — произвольный вектор из $\mathbb{R}^p$, $\varepsilon \sim \mathcal{N}(0, R_\varepsilon)$, $A^*A \succ O$, $R_\varepsilon \succ O$, оценка Гаусса—Маркова

$$\hat{\theta} = \hat{G}Y, \quad \hat{G} = \left(A^* R_\varepsilon^{-1} A\right)^{-1} A^* R_\varepsilon^{-1},$$

минимаксна на классе произвольных оценок:

$$\sup_{\theta\in\mathbb{R}^p} \mathbf{E}_\theta \, |\hat{\theta} - \theta|^2 \leq \sup_{\theta\in\mathbb{R}^p} \mathbf{E}_\theta \, |\widetilde{\theta} - \theta|^2 \quad \forall \widetilde{\theta}.$$

Доказательство. Оцениватель Гаусса—Маркова можно представить в виде предела байесовских оценивателей:

$$\hat{G} = \lim_{d\uparrow\infty} \hat{G}^{(d)}, \quad \hat{G}^{(d)} = \left(A^* R_\varepsilon^{-1} A + d^{-1} I\right)^{-1} A^* R_\varepsilon^{-1},$$

где $\hat{G}^{(d)} Y = \mathbf{E}^{(d)}\{\theta \mid Y\}$ вычисляется в предположении, что $\theta \sim \mathcal{N}(0, d\, I)$ и не зависит от ε.

Минимаксность линейной несмещенной оценки

Вопрос

Можно ли распространить теорему Ходжеса—Лемана на случай обобщенной модели

$$\begin{cases} X = a\theta + \xi, \\ Y = A\theta + \eta, \end{cases} \qquad \theta \in \mathbb{R}^p, \quad \upsilon = \begin{bmatrix} \xi \\ \eta \end{bmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix}\right)$$

без требований о невырожденности матриц A^*A, R_Y и т.д.?

Теорема 2

Наилучшая линейная несмещенная оценка $\hat{X} = \hat{L}Y$ является минимаксной на классе произвольных оценок $\widetilde{X} = \varphi(Y)$, $\varphi \in \mathcal{B}$:

$$\sup_{\theta \in \mathbb{R}^p} \mathfrak{J}_\theta(\hat{L}, \mathbf{P}_\upsilon) \leq \sup_{\theta \in \mathbb{R}^p} \mathfrak{J}_\theta(\varphi, \mathbf{P}_\upsilon) \quad \forall\, \varphi \in \mathcal{B},$$

где $\mathbf{P}_\upsilon = \mathcal{N}(0, R)$.

Минимаксность: достаточное условие

Лемма 4 (Гарантированный риск равен байесовскому риску)

Оценка $\hat{X}$ будет минимаксной, если

$$\boxed{\sup_{\theta\in\mathbb{R}^p} \mathbf{E}_\theta |\hat{X} - X|^2} = \boxed{\sup_{\Pi\in\mathcal{P}_\theta} \mathbf{E}^{(\Pi)} |\hat{X}^{(\Pi)} - X|^2},$$

где Π — априорное распределение из некоторого класса $\mathcal{P}_\theta$,

$$\mathbf{E}^{(\Pi)}\{\dots\} = \int \mathbf{E}_\theta\{\dots\}\,\Pi(d\theta), \qquad \hat{X}^{(\Pi)} = \mathbf{E}^{(\Pi)}\{X \mid Y\}.$$

Положим $\mathcal{P}_\theta = \{\mathcal{N}(0,T)\colon\ T \succeq O\}$. Тогда

$$\boxed{\sup_{\Pi\in\mathcal{P}_\theta} \mathbf{E}^{(\Pi)} |\hat{X}^{(\Pi)} - X|^2 = \sup_{T\succeq O} \min_{L\in\mathcal{L}} J(L,T),}$$

где $\mathcal{L}$ — класс линейных операторов оценивания L,

$$J(L,T) = \operatorname{tr}[(LA-a)T(LA-a)^* + R_X - 2LR_{YX} + LR_Y L^*].$$

Минимаксность на классе всех линейных оценок

В байесовской модели, где $\theta \sim \mathcal{N}(0, T)$, с.к. риск линейной оценки $\widetilde{X} = LY$:

$$J(L, T) = \mathrm{tr}[(LA - a)T(LA - a)^* + R_X - 2LR_{YX} + LR_Y L^*]\,.$$

В исходной модели, где θ — вектор неопределенных параметров, с.к. риск линейной оценки $\widetilde{X} = LY$ ограничен

$$\sup_{\theta \in \mathbb{R}^p} \mathbf{E}_\theta |\widetilde{X} - X|^2 = \sup_{\theta \in \mathbb{R}^p} J(L, \theta\theta^*) = \sup_{T \succeq O} J(L, T) < \infty$$

только если $\widetilde{X} = LY$ — несмещенная оценка, т.е. $LA = a$.

Поэтому с.к. риск наилучшей линейной несмещенной оценки $\hat{X} = \hat{L}Y$ равен:

$$\boxed{\sup_{\theta \in \mathbb{R}^p} \mathbf{E}_\theta |\hat{X} - X|^2 = \min_{L \in \mathcal{L}} \sup_{T \succeq O} J(L, T).}$$

Минимаксность: завершение доказательства

Для доказательства теоремы остается проверить:

$$\inf_{L\in\mathcal{L}} \sup_{T\succeq O} J(L,T) = \sup_{T\succeq O} \inf_{L\in\mathcal{L}} J(L,T), \qquad (*)$$

- функция $J(L,T)$ — выпуклая по L и вогнутая по T,
- переменные L, T пробегают выпуклые замкнутые, но неограниченные множества.

Теорема 3 (несимметричная теорема о минимаксе)

Равенство $(*)$ имеет место, если при любом T рецессивный конус функции $J(\cdot,T)$ образует линейное подпространство.

Рецессивный конус выпуклой функции

состоит из тех векторов, вдоль которых функция не возрастает.

$J(\cdot,T)$ — квадратичная $\Rightarrow$ ее рецессивный конус — это ее ядро. $\square$

Пример 4 (V.D. Vandelinde (1977) IEEE TAC G. Schwarz (1987) JASA)

Пусть совместное распределение оцениваемого вектора X и вектора наблюдений Y принадлежит классу

$$\mathcal{P}(0,R)=\left\{\mathbf{P}_\upsilon:\ \mathbf{E}\upsilon=0,\ \mathbf{cov}\{\upsilon,\upsilon\}=R=\begin{pmatrix} R_X & R_{XY}\\ R_{YX} & R_Y\end{pmatrix}\right\}.$$

В гауссовском случае линейная оценка $\hat{X}=\hat{\Lambda}Y$, $\hat{\Lambda}=R_{XY}R_Y^+$, будет байесовской $\Longrightarrow$ для любого оценивателя $\varphi\in\mathcal{B}$

$$\boxed{\mathfrak{J}(\hat{\Lambda},\widehat{\mathbf{P}}_\upsilon)\le\mathfrak{J}(\varphi,\widehat{\mathbf{P}}_\upsilon)},\quad \text{где}\quad \widehat{\mathbf{P}}_\upsilon=\mathcal{N}(0,R).$$

Но $\hat{\Lambda}$ — уравнивающая стратегия, так как для любого допустимого распределения $\mathbf{P}_\upsilon\in\mathcal{P}(0,R)$

$$\boxed{\mathfrak{J}(\hat{\Lambda},\mathbf{P}_\upsilon)=\mathfrak{J}(\hat{\Lambda},\widehat{\mathbf{P}}_\upsilon)}=\mathrm{tr}\left[R_X-R_{XY}R_Y^+R_{YX}\right].$$

Поэтому <u>линейный оцениватель</u> $\hat{\Lambda}=R_{XY}R_Y^+$ и <u>гауссовское распределение</u> $\widehat{\mathbf{P}}_\upsilon=\mathcal{N}(0,R)$ дают седловую точку с.к. риска:

$$\mathfrak{J}(\hat{\Lambda},\mathbf{P}_\upsilon)\le\mathfrak{J}(\hat{\Lambda},\widehat{\mathbf{P}}_\upsilon)\le\mathfrak{J}(\varphi,\widehat{\mathbf{P}}_\upsilon)\quad\forall\varphi\in\mathcal{B}\quad\forall\mathbf{P}_\upsilon\in\mathcal{P}(0,R).$$

Случай неизвестного распределения

Вопрос

Пусть в модели $X = a\theta + b\upsilon, \quad Y = A\theta + B\upsilon, \quad \theta \in \mathbb{R}^p$ случайный вектор υ имеет любое распределение, такое что

$$\mathbf{E}\,\upsilon = 0, \quad \mathbf{cov}\{\upsilon, \upsilon\} = R, \quad \text{т.е.} \quad \mathbf{P}_\upsilon \in \mathcal{P}(0, R).$$

- Можно ли минимаксную оценку $\hat{X}$ выбрать линейной?
- Что если матрица R априори принадлежит множеству $\mathcal{R}$?

С.к. риск линейной несмещенной оценки $\widetilde{X} = LY$, $L \in \mathcal{L}_0$:

$$\mathfrak{J}_\theta(L, \mathbf{P}_\upsilon) = J(L, R) = \mathrm{tr}[(LB - b)R(LB - b)^*] \ \forall\, \mathbf{P}_\upsilon \in \mathcal{P}(0, R) \ \forall\, \theta.$$

Лемма 5 (Седловая точка в задаче линейного оценивания)

Если $\mathcal{R}$ выпукло и компактно, то существуют линейный несмещенный оцениватель $\hat{L} \in \mathcal{L}_0$ и допустимая ковариационная матрица $\hat{R} \in \mathcal{R}$, образующие седловую точку

$$J(\hat{L}, R) \le J(\hat{L}, \hat{R}) \le J(L, \hat{R}) \quad \forall\, (L, R) \in \mathcal{L}_0 \times \mathcal{R}.$$

Теорема 4

Линейная минимаксная оценка $\hat{X} = \hat{L}Y$ *является одновременно минимаксной на классе произвольных оценок:*

$$\sup_{\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon} \sup_{\theta \in \mathbb{R}^p} \mathfrak{J}_\theta(\hat{L}, \mathbf{P}_\upsilon) \leq \sup_{\mathbf{P}_\upsilon \in \mathcal{P}_\upsilon} \sup_{\theta \in \mathbb{R}^p} \mathfrak{J}_\theta(\varphi, \mathbf{P}_\upsilon) \quad \forall\, \varphi \in \mathcal{B}, \qquad (**)$$

где $\mathcal{P}_\upsilon = \{\mathbf{P}_\upsilon : \mathbf{E}\,\upsilon = 0,\ \mathbf{cov}\{\upsilon, \upsilon\} \in \mathcal{R}\}$.

Доказательство.

Если положить $\overline{\mathfrak{J}}(\varphi, \mathbf{P}_\upsilon) = \sup_{\theta \in \mathbb{R}^p} \mathfrak{J}_\theta(\varphi, \mathbf{P}_\upsilon)$ и $\widehat{\mathbf{P}}_\upsilon = \mathcal{N}(0, \hat{R})$, то «левое неравенство» $J(\hat{L}, R) \leq J(\hat{L}, \hat{R})$, $R \in \mathcal{R}$, принимает вид

$$\overline{\mathfrak{J}}(\hat{L}, \mathbf{P}_\upsilon) \leq \overline{\mathfrak{J}}(\hat{L}, \widehat{\mathbf{P}}_\upsilon) \quad \forall\, \mathbf{P}_\upsilon \in \mathcal{P}_\upsilon.$$

Из теоремы 2 получаем «правое неравенство»

$$\overline{\mathfrak{J}}(\hat{L}, \widehat{\mathbf{P}}_\upsilon) \leq \overline{\mathfrak{J}}(\varphi, \widehat{\mathbf{P}}_\upsilon) \quad \forall\, \varphi \in \mathcal{B}.$$

Поэтому линейный оцениватель $\hat{L}$ удовлетворяет $(**)$. $\square$

Пример 5 (Оценка Кукса—Ольмана)

Если в модели $Y = A\theta + \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, R_\varepsilon)$, $R_\varepsilon \succ O$, вектор параметров θ принадлежит эллипсоиду $\Theta = \{\theta \in \mathbb{R}^p \colon \langle \Sigma\theta, \theta\rangle \le 1\}$, то линейная минимаксная оценка $\hat{x}$ скалярной величины $x = \langle c, \theta\rangle$ имеет явный вид

$$\hat{x} = \langle c, \left(A^* R_\varepsilon^{-1} A + \Sigma\right)^{-1} A^* R_\varepsilon^{-1} Y\rangle.$$

Вопросы

1. Будет ли оценка Кукса—Ольмана минимаксной на классе всех возможных оценок?
2. Будет ли это верно, если вектор ошибок ε имеет произвольное распределение, такое что

$$\mathbf{E}\,\varepsilon = 0, \quad \mathbf{cov}\{\varepsilon, \varepsilon\} = R_\varepsilon, \quad \text{т.е.} \quad \mathbf{P}_\varepsilon \in \mathcal{P}(0, R_\varepsilon)?$$

Линейность минимаксных оценок: контрпримеры

Пример 6 (G. Casella, W.E. Strawderman (1981) Ann. Stat., N.4)

Пусть $Y \sim \mathcal{N}(\theta, 1)$ — наблюдение, $\theta \in \Theta = [-r, r]$, $r < 1{,}05$. Тогда нелинейная оценка $\hat{\theta} = r\,\mathrm{th}(rY)$ — минимаксная, а линейная оценка Кукса—Ольмана $\hat{\theta}^{\mathcal{L}} = \dfrac{r^2 y}{1 + r^2}$ — нет:

$$\sup_{\theta\in\Theta} \mathbf{E}_\theta (\hat{\theta} - \theta)^2 \approx 0.45 < 0.5 = \sup_{\theta\in\Theta} \mathbf{E}_\theta (\hat{\theta}^{\mathcal{L}} - \theta)^2 \quad \text{при} \quad r = 1.$$

Пример 7 (D.L. Donoho (1994) Ann. Stat., N.1)

Пусть в модели $Y = A\theta + \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$, вектор параметров θ принадлежит выпуклому центрально симметричному компакту $\Theta \subset \mathbb{R}^p$. Если $\hat{x}$ — минимаксная, а $\hat{x}^{\mathcal{L}}$ — минимаксная линейная оценка скалярной величины $x = \langle c, \theta \rangle$, то

$$1 < \sup_{\theta\in\Theta} \mathbf{E}_\theta (\hat{x}^{\mathcal{L}} - x)^2 / \sup_{\theta\in\Theta} \mathbf{E}_\theta (\hat{x} - x)^2 < 1{,}25.$$

Линейность минимаксных оценок: пример

Пример 8

Пусть параметр $\theta \in \Theta = [-r, r]$ оценивается по $Y = \theta + \varepsilon$ с ошибкой, имеющей произвольный закон распределения, такой что $\mathbf{E}\varepsilon = 0$, $\mathbf{D}\varepsilon = 1$, т.е. $\mathbf{P}_\varepsilon \in \mathcal{P}(0,1)$.

Тогда линейная оценка $\hat{\theta}^{\mathcal{L}} = \dfrac{r^2 y}{1 + r^2}$ будет минимаксной:

$$\sup_{\mathbf{P}_\varepsilon \in \mathcal{P}(0,1)} \sup_{\theta \in \Theta} \mathbf{E}_\theta (\hat{\theta}^{\mathcal{L}} - \theta)^2 \leq \sup_{\mathbf{P}_\varepsilon \in \mathcal{P}(0,1)} \sup_{\theta \in \Theta} \mathbf{E}_\theta (\widetilde{\theta} - \theta)^2 \quad \forall\, \widetilde{\theta},$$

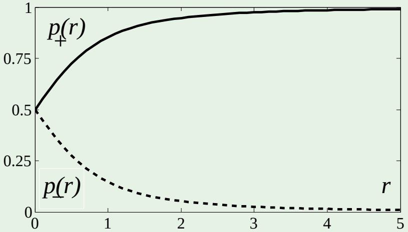
так как $\hat{\theta}^{\mathcal{L}}$ — байесовская оценка, соответствующая наименее благоприятному распределению $\widehat{\mathbf{P}}$:

$$\widehat{\mathbf{P}}(\theta = \pm r) = 1/2,$$

$$\widehat{\mathbf{P}}\left(Y \in \left\{\pm\sqrt{1 + r^2}\right\}\right) = 1,$$

$$\widehat{\mathbf{P}}(Y = \sqrt{1 + r^2} \mid \theta = \pm r) =$$

$$= p_\pm(r) = \frac{1}{(\sqrt{1 + r^2} \mp r)^2 + 1}.$$



Обобщенная модель с ограниченными параметрами

Неопределенно-стохастическая модель наблюдения

$$X = a\theta + b\upsilon, \qquad Y = A\theta + B\upsilon, \qquad \theta \in \mathbb{R}^p,$$

где вектор υ имеет произвольное распределение с частично известными моментными характеристиками:

$$u = \mathbf{E}\,\upsilon \in \mathcal{U}, \qquad R = \mathbf{cov}\{\upsilon, \upsilon\} \in \mathcal{R}.$$

Рассматриваются линейные оценки $\widetilde{X} = LY$, такие что

$$L \in \mathcal{L}_0, \quad \text{т.е.} \quad LA = a.$$

С.к. риск не зависит от вектора неопределенных параметров θ:

$$J(L, K) = \mathrm{tr}[(LB - b)K(LB - b)^*], \quad K \in \mathcal{K},$$

$\mathcal{K}$ — множество допустимых матриц вторых моментов $\mathbf{E}\{\upsilon\upsilon^*\}$

$$\mathcal{K} = \{uu^* + R\colon\ u \in \mathcal{U},\ R \in \mathcal{R}\}.$$

Метод двойственной оптимизации

Задача минимаксного линейного оценивания

$$\hat{L} \in \operatorname*{arg\,min}_{L \in \mathcal{L}_0} \sup_{K \in \mathcal{K}} J(L, K).$$

Метод двойственной оптимизации

1. Решить задачу оптимального линейного оценивания:
   - найти оптимальный оператор $L_K \in \operatorname*{arg\,min}_{L \in \mathcal{L}_0} J(L, K)$
   - и двойственный функционал $\underline{J}(K) = \min_{L \in \mathcal{L}_0} J(L, K)$.

   Решить двойственную задачу $\hat{K} \in \operatorname*{arg\,max}_{K \in \mathcal{K}} \underline{J}(K)$.

3. Построить минимаксный оператор в виде $\hat{L} = L_{\hat{K}}$.

Вопрос

Всегда ли этот метод приводит к искомому решению?

Пример 9

Пусть $Y_i = c_i\theta + \varepsilon_i$, $\theta \in \mathbb{R}$, $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = 1$, $i = 1, \dots, n$, т.е.

$$\mathbf{cov}\{\varepsilon, \varepsilon\} \in \mathcal{R} = \{R \succeq O\colon\ R_{ii} = 1,\ i = 1, \dots, n\}.$$

Задача минимаксного линейного оценивания:

$$\max_{R \in \mathcal{R}} \mathbf{D}\langle \lambda, Y\rangle = \|\lambda\|_1^2 = \left(\sum_{i=1}^{n} |\lambda_i|\right)^2 \longrightarrow \min_{\lambda\colon\ \langle \lambda, c\rangle = 1}.$$

Пусть $i_0\colon\ |c_{i_0}| > |c_i|\ \forall\, i \neq i_0$, тогда получаем:

- минимаксную линейную оценку
$$\hat{\theta} = \langle \hat{\lambda}, Y\rangle = Y_{i_0}/c_{i_0},$$
- минимаксный риск
$$\hat{J} = \max_{R \in \mathcal{R}} \mathbf{D}\hat{\theta} = c_{i_0}^{-2} = \frac{1}{\max(c_1^2, \dots, c_n^2)}.$$

Пример 9 (продолжение)

Двойственная задача: $\underline{J}(R) = \min\limits_{\lambda:\,\langle\lambda,c\rangle=1} \mathbf{D}\langle\lambda, Y\rangle \longrightarrow \max\limits_{R\in\mathcal{R}}$.

$$\underline{J}(R) = c^+(R - R(QRQ)^+R)(c^+)^* \le c^+R(c^+)^* = \ldots$$

(здесь $Q = I - cc^+$). Положим $\hat{R} = c_{i_0}^{-2}cc^* \Rightarrow \hat{R} \in \mathcal{R} \Rightarrow$

$$\ldots = |c|^{-4}\langle Rc, c\rangle \le |c|^{-4}\langle \hat{R}c, c\rangle = c_{i_0}^{-2} = \hat{J}.$$

$\Longrightarrow \hat{R}$ — наименее благоприятная, но соответствующая оптимальная линейная оценка $\widetilde{\theta}^o$ — неминимаксная:

$$\widetilde{\theta}^o = c^+[I - \hat{R}(Q\hat{R}Q)^+]Y = c^+Y = \frac{c_1Y_1 + \ldots + c_nY_n}{c_1^2 + \ldots + c_n^2},$$

$$\max_{R\in\mathcal{R}} \mathbf{D}\,\widetilde{\theta}^o = \|c^+\|_1^2 = \left(\frac{|c_1| + \ldots + |c_n|}{c_1^2 + \ldots + c_n^2}\right)^2 > \frac{1}{\max(c_1^2, \ldots, c_n^2)} = \hat{J}.$$

Если $c_k = \frac{1}{k}$, то $\max\limits_{R\in\mathcal{R}} \sqrt{\mathbf{D}\,\widetilde{\theta}^o} \approx \frac{\ln n + \gamma}{\pi^2/6} \gg 1 = \sqrt{\hat{J}}$ при $n \sim \infty$.

Метод двойственной оптимизации: обоснование

Неопределенно-стохастическая модель наблюдения

$$X = a\theta + b\upsilon, \quad Y = A\theta + B\upsilon, \quad \theta \in \mathbb{R}^p, \quad K = \mathbf{E}\{\upsilon\upsilon^*\} \in \mathcal{K}.$$

С.к. риск линейной несмещенной оценки $\widetilde{X} = LY$, $L \in \mathcal{L}_0$:

$$\mathbf{E}\,|\widetilde{X} - X|^2 = J(L, K) = \mathrm{tr}[(LB - b)K(LB - b)^*] \quad \text{при} \quad K = \mathbf{E}\{\upsilon\upsilon^*\}.$$

Теорема 5

Если $\mathcal{K}$ — компакт и в наименее благоприятном случае

$$\hat{K} \in \arg\max_{K \in \mathrm{co}[\mathcal{K}]} \underline{J}(K), \quad \underline{J}(K) = \min_{L \in \mathcal{L}_0} J(L, K),$$

матрица $\mathbf{E}\{YY^*\}$ невырожденная, т.е. $B\hat{K}B^* \succ O$,
то соответствующий оптимальный оператор оценивания $L_{\hat{K}}$ будет минимаксным:

$$L_{\hat{K}} = \arg\min_{L \in \mathcal{L}_0} J(L, \hat{K}) \quad \Longrightarrow \quad L_{\hat{K}} \in \arg\min_{L \in \mathcal{L}_0} \max_{K \in \mathcal{K}} J(L, K).$$

Кроме того, пара $L_{\hat{K}}, \hat{K}$ — седловая точка $J(\cdot)$ на $\mathcal{L}_0 \times \mathrm{co}[\mathcal{K}]$.

Пусть модель наблюдения

$$X = a\theta + b\upsilon, \quad Y = A\theta + B\upsilon, \quad \theta \in \mathbb{R}^p, \quad K = \mathbf{E}\{\upsilon\upsilon^*\} \in \mathcal{K}.$$

регулярна, т.е. $K_Y = BKB^* \succ O \quad \forall\, K \in \mathcal{K}$.

- Оптимальный оператор оценивания:

$$L_K = \Lambda_K + (a - \Lambda_K A) G_K,$$

  - $G_K = \left(A^* K_Y^{-1} A\right)^+ A^* K_Y^{-1}$ — оцениватель Гаусса—Маркова,
  - $\Lambda_K = K_{XY} K_Y^{-1}$ — оцениватель из теоремы о нормальной корреляции.

Двойственный функционал:

$$\underline{J}(K) = \operatorname{tr}\left[K_X - K_{XY}K_Y^{-1}K_{YX} + (a - \Lambda_K A)\left(A^* K_Y^{-1} A\right)^+ (a - \Lambda_K A)^*\right],$$

где $K_X = bKb^*$, $K_{XY} = bKB^*$, $K_Y = BKB^*$.

Двойственная задача: $\hat{K} \in \arg\max\limits_{K \in \mathcal{K}} \underline{J}(K)$, $\underline{J}(K) = \min\limits_{L \in \mathcal{L}_0} J(L, K)$.

Алгоритм 1 (Метод условного градиента)

0. Положить $s := 0$ и выбрать $K^0 \in \mathcal{K}$.
1. Определить оптимальный оператор $L^s = \arg\min\limits_{L \in \mathcal{L}_0} J(L, K^s)$.
2. Найти наихудшую матрицу $\widetilde{K}^s \in \arg\max\limits_{K \in \mathcal{K}} J(L^s, K)$.
3. Вычислить $\delta^s = \underline{J}'(K^s)(\widetilde{K}^s - K^s) = J(L^s, \widetilde{K}^s - K^s) \geq 0$.
4. Если $\delta^s = 0$, то $\boxed{\hat{L} := L^s,\ \hat{K} := K^s \text{ — седловая точка}}$.
5. Найти величину шага $t^s \in \arg\max\limits_{t \in [0,1]} \underline{J}(K^s + t(\widetilde{K}^s - K^s))$;
6. Положить $K^{s+1} := K^s + t^s(\widetilde{K}^s - K^s)$, $s := s + 1$ и перейти к шагу 1.

Сходимость алгоритма решения двойственной задачи

Теорема 6

Пусть множество $\mathcal{K}$ выпукло и компактно и выполнено условие регулярности $K_Y = BKB^* \succ O \quad \forall K \in \mathcal{K}$.

1. *$\{L^s\}$ сходится к минимаксному оценивателю $\hat{L}$, а $\{K^s\}$ — к множеству решений двойственной задачи.*
2. *Оценка погрешности в прямой и двойственной задачах:*

$$\max_{K \in \mathcal{K}} J(L^s, K) - \hat{J} \leq \delta^s, \qquad \Delta^s = \hat{J} - \underline{J}(K^s) \leq \delta^s,$$

($\hat{J}$ — минимаксный риск).

3. *Априорная оценка погрешности:*

$$\Delta^s \leq \left(\frac{1}{\Delta^0} + \frac{s}{2 \max\{HD^2, \Delta^0\}} \right)^{-1} = O(1/s),$$

$$\|L^s - \hat{L}\|_2 \leq \left(\Delta^s / \min_{K \in \mathcal{K}} \sigma_{\min}[BKB^*] \right)^{1/2} = O(1/\sqrt{s}),$$

(H — константа Липшица $\nabla \underline{J}(\cdot)$ на $\mathcal{K}$, D — диаметр $\mathcal{K}$).

Задача полуопределенного программирования (SDP)

$$\langle c, x\rangle \to \min_{x\in\mathbb{R}^N} \quad \text{при ограничении} \quad \mathcal{A}(x) \succeq O,$$

заданы вектор $c \in \mathbb{R}^N$ и аффинно-линейное отображение $\mathcal{A}$ из $\mathbb{R}^N$ в пространство симметричным $M \times M$ матриц.

Для решения SDP есть пакеты, реализованные в MATLAB:

- SeDuMi: J.F. Sturm (1999) Optim. Methods Software. N. 12.
- CVX: M. Grant, S. Boyd (2011) `http://cvxr.com/cvx/download`

Лемма Шура

Для любых $\mathrm{A} \succeq O$ *и* $\mathrm{B} \succ O$ *справедливо утверждение*

$$\mathrm{A} - \Gamma^*\mathrm{B}^{-1}\Gamma \succeq O \quad\Longleftrightarrow\quad \begin{bmatrix} \mathrm{A} & \Gamma^* \\ \Gamma & \mathrm{B} \end{bmatrix} \succeq O.$$

Пример 10 (Применение SDP к стохастической регрессии)

Пусть ковариационная матрица R оцениваемого вектора X и вектора наблюдений Y

$$R = \mathbf{cov}\{\upsilon, \upsilon\} = \begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix}, \quad \text{где} \quad \upsilon = \begin{pmatrix} X \\ Y \end{pmatrix},$$

принадлежит множеству

$$\mathcal{R} = \{R \succeq O \colon \ |R_{ij} - R^o_{ij}| \le \Delta_{ij} \ \forall\, i, j\}.$$

где R^o — некоторая опорная матрица.

Двойственная задача: $\quad \operatorname{tr}\left[R_X - R_{XY} R_Y^{-1} R_{YX}\right] \longrightarrow \max\limits_{R \in \mathcal{R}}.$

Вводим новую переменную T и ограничение на нее:

$$T \preceq R_X - R_{XY} R_Y^{-1} R_{YX}, \quad \text{т.е.} \quad R_X - T - R_{XY} R_Y^{-1} R_{YX} \succeq O.$$

Тогда по лемме Шура двойственная задача — это SDP:

$$\operatorname{tr}[T] \longrightarrow \max_R \colon \ \begin{pmatrix} R_X - T & R_{XY} \\ R_{YX} & R_Y \end{pmatrix} \succeq O, \ |R_{ij} - R^o_{ij}| \le \Delta_{ij} \ \forall\, i, j.$$

Вопросы и проблемы

1. Исследовать проблему линейности минимаксных оценок в моделях, содержащих ограниченные параметры и случайные ошибки с неизвестным распределением.
2. Как быть, если критерий не среднеквадратичный, а вероятностный, квантильный и т.д.?
3. Поиск менее консервативных минимаксных постановок.
4. Как применить SDP к решению задачи минимаксного оценивания в неопределенно-стохастических моделях с неограниченными параметрами?
5. Что даст SDP в моделях наблюдения с ограниченными параметрами?

Спасибо за внимание!