Семинар 1

Оценки. Несмещенность. Квадратичное отклонение

1.1 Базовая часть

1.1.1 Точечное оценивание

Считается, что вы уже ознакомились с введением, в котором описывалась логика построения статистической модели. Здесь же я непосредственно буду пользоваться этой моделей без дополнительного пояснения.

Пусть $(\mathcal{X}, \mathbb{F}, \{\mathbf{P}_\theta(\cdot), \theta \in \Theta\})$ – статистическое пространство, $\Theta \subseteq \mathbb{R}^k$, $\mathcal{X} \subseteq \mathbb{R}^m$, k, m – некоторые константы.

Мы рассматриваем случайный вектор $\vec{X}$ со значениями в $\mathcal{X}$, имеющий распределение $\mathbf{P}_\theta$ при некотором $\theta \in \Theta$. Задача точечного оценивания предполагает построение измеримой (борелевской) функции $\widehat{\theta}$ из $\mathcal{X}$ в $\mathbb{R}^k$, которую мы называем оценкой. Мы ожидаем, что $\widehat{\theta}(\vec{X})$ в некотором смысле близка к функции $h(\theta)$, которую мы хотим оценить.

Какими свойствами должна обладать наша оценка? Что означает "близка"? Мы рассмотрим несколько возможных вариантов и подходов.

Отметим, что приведенные выше соображения относятся к параметрическому случаю, с которым мы преимущественно будем работать. В непараметрическом случае мы рассматриваем пространство $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, где $\mathcal{P}$ – непараметризованное множество мер. При этом мы также строим оценку $\widehat{\theta}$, но уже не для функции $h(\theta)$, а для параметра распределения $h(P)$, где $h : \mathcal{P} \to \mathbb{R}^k$ – некоторое отображение из множества мер в множество векторов.

1.1.2 Точечное оценивание

Итак, пусть у нас есть выборка $\vec{X}$ из распределения $\mathbf{P}_\theta$. Мы бы хотели построить функцию $\widehat{\theta}(\vec{X})$, которая бы хорошо оценивала θ. Для начала стоит задаться хрестоматийным вопросом — "Что такое хорошо?"

Определение 1. Оценка $\widehat{\theta}$ называется несмещенной для функции $g(\theta)$, если $\mathbf{E}_\theta\widehat{\theta}(X) = g(\theta)$ при всех θ.

Определение 2. Смещением оценки $\widehat{\theta}$ называют величину $\mathbf{E}_\theta\widehat{\theta}(X) - g(\theta)$.

Здесь и далее $\mathbf{E}_\theta$ обозначает математическое ожидание по мере $\mathbf{P}_\theta$.

Будем говорить, что оценка несмещенная без указания функции, если подразумевается, что $g(\theta) = \theta$.

Аналогичным образом в непараметрической модели оценка называется несмещенной для параметра распределения $h(P)$, если

$$\mathbf{E}_P\widehat{\theta}(\vec{X}) = h(P), \quad P \in \mathcal{P}.$$

Здесь $\mathbf{E}_P$ – математическое ожидание по мере P.

В дальнейшем мы будем рассматривать случай, когда выборка представляет собой вектор из н.о.р. величин $X_1, \ldots, X_n$. Соответственно, многомерное распределение $\mathbf{P}_\theta$ мы будем задавать одномерным распределением X_1.

Пример 1. Предположим, что на основе неких измерений в течение дня, мы делаем вывод о необходимых закупках товара в магазин. Несмещенность, в силу закона больших чисел, гарантирует, что средняя закупка за N дней будет стремиться с ростом N к искомому параметру. Смещенные оценки будут иметь систематический снос, то мы есть будем систематически недокупать (перекупать) товар.

С другой стороны, сама по себе несмещенность не обеспечивает "разумности" оценки.

Пример 2. Рассмотрим схему Бернулли с неизвестным параметром θ и рассмотрим оценку $\overline{X} = (X_1 + ... + X_n)/n$. Иначе говоря, чтобы оценить вероятность успеха, мы делим число успехов на число испытаний, что вполне естественно. В силу линейности математического ожидания, эта оценка несмещенная. При этом можно взять оценку X_1. Оценка несмещена, поскольку $\mathbf{E}_\theta X_1 = \theta$, но явно не слишком-то удобна для использования. Как минимум, она принимает только значения 0 и 1 и зависит лишь от первого элемента выборки.

Пример 3. Собственно, даже в общей непараметрической модели оценка $\overline{X}$ будет несмещенной для $\mathbf{E}_P X$, где $\mathcal{P} = \{P : \mathbf{E}_P X < \infty\}$. Действительно,

$$\mathbf{E}_P \overline{X} = \frac{1}{n} \mathbf{E}_P (X_1 + \ldots + X_n) = \mathbf{E}_P X_1.$$

Поэтому в любой модели эта оценка является несмещенной для математического ожидания. Такая оценка широко используется и носит название *выборочного среднего*.

Будем обозначать $\overline{f(X)}$ величину $n^{-1} \sum_{i=1}^n f(X_i)$. Аналогичным образом эта оценка несмещена для $\mathbf{E}_P f(X)$, если это математическое ожидание конечно.

Пример 4. Поскольку

$$h(P) = \mathbf{P}(X \leq x) = \mathbf{E}_P I_{X \leq x},$$

то в рамках предыдущего примера мы можем использовать несмещенную оценку

$$\widehat{F}_n(x) = \overline{I_{X \leq x}} = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x}$$

для оценки ф.р. наших наблюдений. Эта оценка играет большую роль в непараметрической статистике и называется эмпирической функцией распределения.

Пример 5. А как оценить дисперсию? Для этого используют оценку, которую называют *выборочной дисперсией*:

$$S^2 := \overline{(X - \overline{X})^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2 = \overline{X^2} - \overline{X}^2.$$

Эта оценка не будет несмещенной, но ее легко исправить так, что она станет несмещенной.

1.1.3 Построение несмещенных оценок

Отметим существенную проблему, усложняющую построение несмещенных оценок — если $\mathbf{E}\widehat{\theta}(X_1, ..., X_n) = g(\theta)$, то $\mathbf{E}h(\widehat{\theta}(X_1, ..., X_n))$, вообще говоря, не равно $h(g(\theta))$. Более того, для строго выпуклых h и невырожденных (то есть непостоянных) случайных величин Y верно неравенство Иенсена

$$\mathbf{E}h(Y) > h(\mathbf{E}Y).$$

Тем самым, построение в каждой модели несмещенных оценок требует отдельных расчетов для различных функций $g(\theta)$. Другие свойства, которые мы пройдем на следующих занятиях, зачастую более просты для исследования. Более того, их может и вовсе не оказаться.

Пример 6. Существует ли несмещенная оценка для $1/\theta$ на основе одного наблюдения $X \sim R[0,\theta]$, $\theta > 0$? Пусть да. Тогда

$$\mathbf{E}_\theta \widehat{\theta}(X) = \int_0^\theta \frac{1}{\theta}\widehat{\theta}(x)dx = \frac{1}{\theta}.$$

Значит, при всех θ получаем

$$\int_0^\theta \widehat{\theta}(x)dx = 1$$

при всех $\theta > 0$. Но левый интеграл стремится к нулю при $\theta \to 0$, а правый нет. Противоречие показывает, что несмещенных оценок не существует.

Пример 7. Отметим, что если в той же задаче рассмотреть $\Theta = (1,\infty)$, то несмещенной оценкой будет $I_{X\in[0,1]}$. Это важное замечание – существование несмещенной оценки существенно зависит от области изменения параметра.

Зачастую можно искать несмещенную оценку подбором, пользуясь известными оценками.

Пример 8. Построить несмещенную оценку для θ^{-2} на основе $X_i \sim \exp(\theta)$.
Математическое ожидание $\mathbf{E}_\theta X_1 = \theta^{-1}$, поэтому можно предположить, что $\overline{X}$ хорошая оценка для θ^{-1}, а $\overline{X}^2$ для θ^{-2}. Но если $\mathbf{E}_\theta \overline{X} = \theta^{-1}$, то $\mathbf{E}_\theta \overline{X}^2 > \theta^{-2}$, поскольку $\mathbf{E}Y^2 > (\mathbf{E}Y)^2$ для любой непостоянной Y. Сможем ли мы исправить смещение $\overline{X}^2$? Заметим, что $\sum_{i=1}^n X_i \sim \Gamma\left(n, \frac{1}{\theta}\right)$, поэтому

$$\mathbf{E}\overline{X}^2 = \frac{1}{n^2\Gamma(n)}\int_0^\infty x^2 x^{n-1} e^{-\frac{x}{\theta}}dx = \frac{\Gamma(n+2)}{\Gamma(n)n^2}\theta^2 = \frac{(n+1)n}{n^2}\theta^2 = \frac{n+1}{n}\theta^2.$$

Пример 9. Построим несмещенную оценку для $e^{-\theta}$ для выборки с распределение $Poiss(\theta)$.
Найдем оценку, зависящую только от одного наблюдения, $\widehat{\theta}(X_1) = h(X_1)$. Тогда

$$\mathbf{E}_\theta h(X_1) = \sum_{k=0}^\infty h(k)\frac{e^{-\theta}\theta^k}{k!} = e^{-\theta},$$

то есть

$$\sum_{k=0}^\infty \frac{h(k)\theta^k}{k!} = 1.$$

Легко угадать решение — $h(k) = I_{k=0}$. Задумавшись, вы легко поймете почему это единственное решение и как найти его, не угадывая с помощью ряда Тейлора. Формально оценка $I_{X_1=0}$ — уже решение задачи, но давайте построим более качественную оценку

$$\overline{I_{X=0}} = \frac{1}{n}\sum_{i=1}^n I_{X_i=0},$$

у которой то же математическое ожидание, но которая гораздо разумнее как оценка (как минимум, она зависит от всех наблюдений, а не только от первого).

1.1.4 Квадратичное отклонение

Измерять отклонение оценки от оцениваемой функции $g(\theta)$ естественно оценивать квадратичным отклонением

$$\mathbf{E}_\theta(\widehat{\theta}\left(\vec{X}) - g(\theta)\right)^2.$$

Для несмещенных оценок данная величина есть

$$\mathbf{D}_\theta\widehat{\theta}(\vec{X}).$$

Для произвольных оценок нетрудно вывести аналогичный факт:

$$\mathbf{E}_\theta\left(\widehat{\theta}(\vec{X}-g(\theta)\right)^2=\mathbf{D}_\theta\widehat{\theta}(\vec{X})+\left(\mathbf{E}_\theta\widehat{\theta}(\vec{X})-g(\theta)\right)^2$$

Таким образом, квадратичное отклонение оценки есть сумма квадрата смещения и дисперсии оценки.

Пример 10. Оценка $\overline{X}$ как оценка $\mathbf{E}_P X$ имеет квадратичное отклонение

$$\mathbf{D}_P\overline{X}=\frac{1}{n^2}\mathbf{D}_P\left(X_1+\cdots+\mathbf{D}_P X_n\right)=\frac{1}{n}\mathbf{D}_P X_1.$$

Оценка $(X_1+\cdots+X_n+1)/(n+2)$ имеет квадратичное отклонение

$$\mathbf{D}_P\frac{X_1+\cdots+X_n}{n+2}+\left(\mathbf{E}_P\frac{X_1+\cdots+X_n+1}{n+2}-\mathbf{E}_P X_1\right)^2=\frac{n\mathbf{D}_P X_1}{(n+2)^2}+\frac{1}{(n+2)^2}(1-2\mathbf{E}_P X_1)^2.$$

Определение 3. Оценка $\widehat{\theta}$ называется оптимальной в классе $\mathcal{E}$ оценок, если она несмещенная и

$$\mathbf{D}_\theta\widehat{\theta}(\vec{X})\le\mathbf{D}_\theta\widetilde{\theta}(\vec{X}),\quad\widetilde{\theta}(\vec{X})\in E,\theta\in\Theta.$$

Оптимальная оценка имеет наименьшую дисперсию среди всех оценок из $\mathcal{E}$. В дальнейшем мы научимся строить оптимальные оценки в классе всех несмещенных оценок, пока же мы можем строить оптимальные оценки в небольших классах оценок, напрямую сравнивая дисперсии.

Пример 11. Рассмотрим оценки вида $T_{a,b}=a\overline{X}+b\overline{Y}$, где X_i, $i\le n$,– н.о.р. $\mathcal{N}(\theta,\sigma^2)$, Y_i, $i\le m$, – н.о.р. $\mathcal{N}(\theta,2\sigma^2)$. Найдем в этом классе оптимальную оценку θ. Поскольку

$$\mathbf{E}_{\theta,\sigma^2}T_{a,b}=(a+b)\theta,\quad\mathbf{D}_{\theta,\sigma^2}T_{a,b}=\left(\frac{a^2}{n}+\frac{2b^2}{m}\right)\sigma^2.$$

Таким образом, несмещенная оценка соответствует $a+b=1$, при этом условии мы должны минимизировать $a^2/n+2b^2/m$. Составляя функцию Лагранжа, получаем

$$L=\frac{a^2}{n}+\frac{2b^2}{m}+\lambda(a+b-1)\to\min,\quad\frac{\partial L}{\partial a}:\ \frac{2a}{n}+\lambda=0,$$
$$\frac{\partial L}{\partial b}:\ \frac{4b}{n}+\lambda=0,\quad\frac{\partial L}{\partial\lambda}:(a+b-1)=0.$$

Отсюда $a=2/3$, $b=1/3$ дают оптимальную оценку. Строго говоря, требуется проверить, что это действительно минимум, однако других кандидатов на роль наименьшего значения не наблюдается, поскольку на границе функция стремится к бесконечности.

1.2 Факультатив

1.2.1 Уравнение несмещенности

Вопрос поиска (и существования) несмещенной оценки для заданной функции сводится к наличию решения $f(x_1,...,x_n)$ у уравнения несмещенности: равенстве при всех θ функций

$$\sum_{x_1,\dots,x_n}f_\theta(x_1)...f_\theta(x_n)f(x_1,...,x_n)=g(\theta)$$

в дискретном случае ($f_\theta(x)=\mathbf{P}_\theta(X_1=x)$) или

$$\int_{\mathbb{R}^n}f_\theta(x_1)...f_\theta(x_n)f(x_1,...,x_n)dx_1...dx_n=g(\theta)$$

в абсолютно-непрерывном случае ($f_\theta(x)$ — плотность). Аналогичное уравнение составляется в случае, если мы хотим построить оценку, являющуюся функцией от заданной статистики T, в этом случае мы будем умножать $f(t)$ на плотность или распределение статистики T.

Решение такого рода функциональных уравнений оказывается достаточно сложным фактом, однако, зачастую мы можем основываться на разложении функции в ряд Тейлора или поиске ее обратного преобразования Лапласа. Может быть полезной также теорема единственности:

Теорема 1 (теорема единственности для преобразования Лапласа)**.** *Пусть*

$$\int_0^\infty f_1(x)e^{-\lambda x}dx = \int_0^\infty f_2(x)e^{-\lambda x}dx$$

при всех $\lambda > 0$*. Тогда* $f_1(x) = f_2(x)$ *п.в.*

Пример 12. Пусть $X \sim \exp(\theta)$. Тогда для любой функции $a(\theta)$ существует не более одной несмещенной оценки от одного наблюдения. Действительно, заметим, что

$$\mathbf{E}f(X) = \int_0^\infty f(x)\theta e^{-\theta x}dx = a(\theta),$$

откуда f есть обратное преобразование Лапласа от $\theta^{-1}a(\theta)$, которое либо существует и единственно, либо не существует. Например, для $a(\theta) = 1/\theta$ подходящей функцией будет $f(x) = x$, причем в силу предыдущих утверждений она единственна (с точностью до изменения на множестве меры 0).

1.2.2 Общий непараметрический подход к оцениванию

Как построить оценку в непараметрическом случае? Для этого представим оцениваемый параметр в виде $h(F)$, где F – ф.р. наблюдений. Тогда можно использовать так называемую естественную оценку $h(\widehat{F}_n)$, где $\widehat{F}_n$ – ЭФР.

Скажем, для оценки $\mathbf{E}X = \int_{-\infty}^{\infty} xdF(x)$ мы используем

$$\int_{-\infty}^{\infty} xd\widehat{F}_n(x; x_1, \dots, x_n) = \sum_{i=1}^n x_i \cdot \frac{1}{n} = \overline{x},$$

где мы воспользовались тем, что $\widehat{F}_n$ как функция x есть ф.р. дискретной случайной величины, равновероятно принимающей значения $x_1, \dots, x_n$.

Аналогичным образом для дисперсии естественной оценкой будет S^2.

Можно использовать этот подход и в параметрическом случае, однако, здесь будет больше свободы для задания θ в виде $h(F)$. Так у $\mathcal{N}(\theta, 1)$ параметр среднего является и математическим ожиданием, и, например, медианой распределение, что даст различные оценки.

В достаточно широких условиях можно показать, что естественные оценки обладают рядом сильных качеств. К сожалению, несмещенность не входит в их число.