

Лекция 1: Точечные оценки параметров распределения.

Выборка — числа $x_1, \dots, x_n$ — наблюдения за процессом с течением времени.

Допущение: $x_1, \dots, x_n$ — реализации некоторой случайной величины $\xi_1, \dots, \xi_n : \Omega \rightarrow \mathbb{R}$:

$$\xi_1(\omega) = x_1, \dots, \xi_n(\omega) = x_n.$$

Допущение: ξ_i независимы в совокупности, одинаково распределены.

Пусть функция распределения F_{ξ_i} , породившая выборку, известна с точностью до параметра θ :

$$F_{\xi_i} \in \{F_\theta\}, \quad \theta \in \Theta \subset \mathbb{R}^m$$

ПРИМЕР: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Известно: $\xi_i \sim N(E(\xi), \sigma^2)$

Неизвестны: среднее $\theta_1 = E(\xi)$, дисперсия $\theta_2^2 = \sigma^2$: $\Theta = \mathbb{R} \times \mathbb{R}_+$.

ТОЧЕЧНАЯ ОЦЕНКА

Любая (измеримая) функция от элементов выборки, принимающая значения в параметрическом множестве Θ .

ПРИМЕР: МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

Интуитивная оценка мат. ожидания для величин $x_1, \dots, x_n$ — среднее арифметическое:

$$\hat{\theta}(x_1, \dots, x_n) = \bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$$

Замечание: оценка — функция от разного количества аргументов (любой выборки). Поэтому оценка определяется от **случайной величины** $\hat{\theta} = \hat{\theta}(\xi_1, \dots, \xi_n)$.

Что означает, что оценка хорошая? Почему одни оценки плохие, другие — лучше?

ХАРАКТЕРИСТИКИ ОЦЕНКИ

1. Несмещённость

Несмещённая оценка — такая оценка, что в среднем она равна величине: $E[\hat{\theta}_n] = \theta$

2. Состоятельность

Состоятельная оценка: $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{P} \theta$

Сильно состоятельная оценка: $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{n.н.} \theta$

СХОДИМОСТЬ ПО ВЕРОЯТНОСТИ

$$\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{P} \theta \quad \left[\forall \varepsilon > 0 \quad P(|\hat{\theta}_n - \theta| > \varepsilon) \rightarrow_{n \rightarrow \infty} 0 \right]$$

Смысл: чем больше размер выборки, тем меньше вероятность ошибиться на заранее выбранное ε .

≡ ПРИМЕР: СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Пусть надо найти мат. ожидание θ_1 нормального распределения $N(\theta_1, \theta_2^2)$

Зададим оценку как $\hat{\theta}_n = \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$.

Посчитаем в лоб мат. ожидание: $E[\hat{\theta}_n] = \frac{1}{n} \sum_{i=1}^n E[\xi_i] = \theta_1$

Значит, оценка несмещённая.

Оценка так же является состоятельной: по закону больших чисел, $\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{P} E[\xi_i] = \theta_1$.

≡ ПРИМЕР: ДИСПЕРСИЯ

Пусть известно распределение с точностью до дисперсии $\xi_i \sim N(a, \theta_2^2)$.

Оценка $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n (\xi_i - a)^2$ несмещённая и состоятельна.

Пусть теперь снова $\xi_i \sim N(\theta_1, \theta_2^2)$.

Сделаем оценку дисперсии (интуитивно): $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \left(\xi_i - \frac{1}{n} \sum_{j=1}^n \xi_j \right)^2$. Насколько она плоха?

$E[\hat{\theta}_n] = \frac{1}{n} \sum_{i=1}^n \left(E[\xi_i^2] - \frac{2}{n} ((n-1)E^2[\xi_i] + E[\xi_i^2]) + \frac{n}{n^2} E[\xi_i^2] + \frac{n(n-1)}{n^2} E^2[\xi_i] \right) = \frac{1}{n} \sum_{i=1}^n \left(E[\xi_i^2] \left(\frac{n-1}{n} \right) + E^2[\xi_i] \right)$

$E[\hat{\theta}_n] = \frac{n-1}{n} (E[\xi_i^2] - E^2[\xi_i]) = \frac{n-1}{n} \theta_2^2$

Получили асимптотически несмещённую оценку, пусть состоятельную. Чтобы оценка получилась несмещённой, нужно делить на $(n-1)$:

$$\hat{\theta}_n = \frac{1}{n-1} \sum_{i=1}^n \left(\xi_i - \frac{1}{n} \sum_{j=1}^n \xi_j \right)^2.$$

≡ ПРИМЕР, ГДЕ НЕ СУЩЕСТВУЕТ НЕСМЕЩЁННОЙ ОЦЕНКИ

Пусть случайные величины биномиально распределены $\xi_i \sim \text{Binom}(1, p)$ (1 раз бросаем монету, p – вероятность успеха)

Реализации $x_1, \dots, x_n$ – нули и единицы ($\text{Binom}(k, p)$ – кол-во успехов при k бросаний монетки).

Оценим $\theta = \ln(p)$ – логарифм вероятности успеха. Предположим, существует несмещённая оценка $\hat{\theta}_n = \hat{\theta}_n(\xi_1, \dots, \xi_n) \in \mathbb{R}$.

Тогда

$\ln p = E[\hat{\theta}_n] = \sum_{x_1=0}^1 \dots \sum_{x_n=0}^1 \hat{\theta}_n(\xi_1, \dots, \xi_n) P(\xi_1 = x_1, \dots, \xi_n = x_n) = \sum_{x_1=0}^1 \dots \sum_{x_n=0}^1 \hat{\theta}_n(\xi_1, \dots, \xi_n) P(\xi_1 =$

Заметим, что $P(\xi_i = x_i) = p^{x_i} (1-p)^{1-x_i}$

$E[\hat{\theta}_n] = \sum_{x_1=0}^1 \dots \sum_{x_n=0}^1 \hat{\theta}_n(\xi_1, \dots, \xi_n) p^{\sum x_i} (1-p)^{n-\sum x_i} \in \mathbb{R}_+$

Однако $p \in [0, 1] \rightarrow \ln p \in \mathbb{R}_-$

Лекция 2: Функции риска. Эмпирическая функция распределения.

Пусть есть 2 оценки: $\hat{\theta}_n^1, \hat{\theta}_n^2$, которые мы предположили. Хороши ли они? Как их сравнивать?

Найдём разность $\hat{\theta}_n - \theta$. Что мы заплатим, если равенство не равно 0.

❗ ФУНКЦИЯ ШТРАФА (ПОТЕРЬ)

Функция $u()$, удовлетворяющая условиям:

1. $u(0) = 0$
2. симметричность $u(x) = u(-x)$

3. МОНОТОННОСТЬ

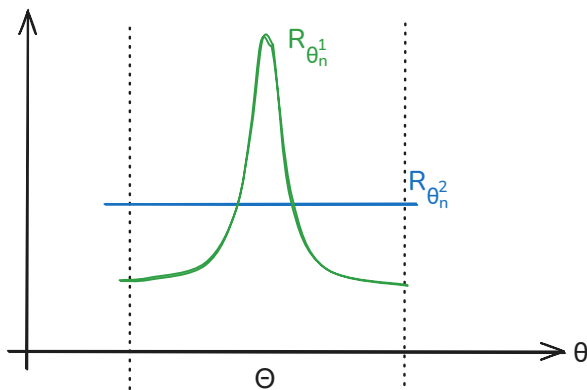
Примеры: парабола $u(x) = x^2$, модуль $u(x) = |x|$.

Функция риска

$$R_{\hat{\theta}_n}(\theta) = E[u(\hat{\theta}_n - \theta)]$$

Замечание: если оценка несмещённая ($\theta = E[\hat{\theta}_n]$), а функция риска квадратичная ($u(x) = x^2$), то функция риска – дисперсия $R_{\hat{\theta}_n}(\theta) = D[\hat{\theta}_n]$

ПРИМЕР



Пусть есть задача минимизировать функцию риска. Здесь $R_{\hat{\theta}_1}(\theta)$ и $R_{\hat{\theta}_2}(\theta)$ ведут себя по-разному: разные площадь под графиком и экстремумы.

Для задачи с большими рисками может лучше подойти $R_{\hat{\theta}_1}(\theta)$.

Подходы выбора функций риска

1. Интегральный (байесовский) подход

$$\int_{\Theta} R_{\hat{\theta}_n}(\theta) d\theta \rightarrow \min_{\hat{\theta}_n}$$

Байесовский – в случае, когда есть априорное распределение

2. Минимаксный подход

$$\max_{\theta \in \Theta} R_{\hat{\theta}_n}(\theta) \rightarrow \min_{\hat{\theta}_n}$$

ПРИМЕР

Пусть у нас 1 параметр: $\xi_i \sim N(\theta, \theta^2)$.

Забыв про хорошую оценку мат. ожидания $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$, рассмотрим некий класс оценок $\hat{\theta}_n \in \{\mu_n \sum_{i=1}^n \xi_i\}$

Квадратичная функция риска: $R_{\hat{\theta}_n}(\theta) = E[(\mu_n \sum_{i=1}^n \xi_i - \theta)^2]$

$$R_{\hat{\theta}_n}(\theta) = E[\mu_n^2 (\xi_1^2 + \dots + \xi_n^2 + \sum_{i \neq j} \xi_i \xi_j) - 2\mu_n \theta \sum_{i=1}^n \xi_i + \theta^2] = \mu_n^2 (2n\theta^2 + n(n-1)\theta^2) - 2\mu_n \theta^2 n + \theta^2 = \theta^2 (n(n+1)\mu_n^2 - 2\mu_n n + 1)$$

Получается, функция риска $R_{\hat{\theta}_n}(\theta)$ разбилась на θ^2 и $(n(n+1)\mu_n^2 - 2\mu_n n + 1)$. Чем меньше мы возьмём второй множитель, тем меньше функция риска.

Для этого подходит $\mu_n = \frac{1}{n+1}$ (и в интегральном, и в минимаксном)

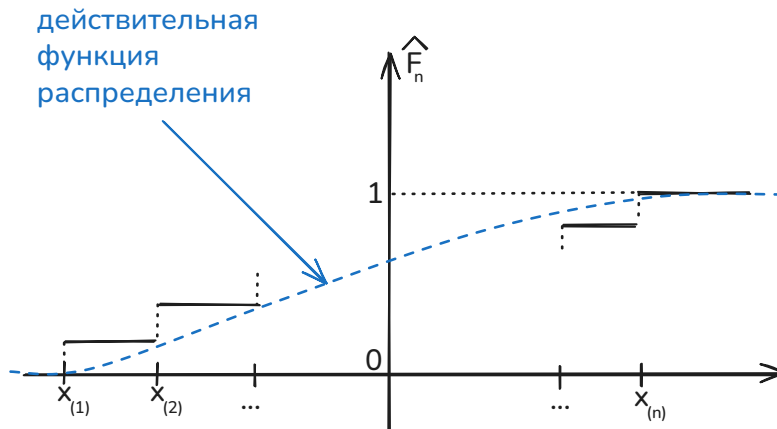
Как по выборке воссоздать функцию распределения $\hat{F}_n$?

ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

$$\hat{F}_n(x; x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n I_{\{\xi_i \leq x\}},$$

где I – индикаторная функция того, что ξ_i не превосходит x .

Упорядочим $x_1, \dots, x_n$ в **вариационный ряд** $x_{(1)}, \dots, x_{(n)}$ ($x_{(k)}$ – k -ая порядковая статистика)



Замечание: пока пишем так, потом $\hat{F}_n(x; \xi_1, \dots, \xi_n)$

УТВЕРЖДЕНИЯ

1. $\forall x \in \mathbb{R} \quad \hat{F}_n(x; \xi_1, \dots, \xi_n)$ – несмещённая оценка для $F_{\xi_i}(x)$
2. $\forall x \in \mathbb{R} \quad \hat{F}_n(x; \xi_1, \dots, \xi_n)$ – состоятельная оценка для $F_{\xi_i}(x)$

ТЕОРЕМА ГЛИВЕНКО-КАНТЕЛЛИ

Для почти любого исхода в пространстве величин, для любой выборки эксперимента не только невязка функции распределения стремится к нулю; к нулю стремится супремум от разности функций.

$$P\left(\sup_{x \in \mathbb{R}} |\hat{F}_n(x; \xi_1, \dots, \xi_n) - F_{\xi_i}(x)| \rightarrow_{n \rightarrow \infty} 0\right) = 1$$

ТЕОРЕМА КОЛМОГОРОВА

Пусть $\sqrt{n}D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |\hat{F}_n(x; \xi_1, \dots, \xi_n) - F_{\xi_i}(x)|$ (D_n – обозначение).

Пусть F_{ξ_i} непрерывна.

Тогда

$$\sqrt{n}D_n \xrightarrow{d}_{n \rightarrow \infty} \eta: \quad F_\eta(y) = \begin{cases} 0, & y \leq 0 \\ \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2 y^2}, & y > 0 \end{cases}$$

Это – функция распределения Колмогорова.

Лекция 3: Асимптотическая нормальность

АСИМПТОТИЧЕСКИ НОРМАЛЬНАЯ ОЦЕНКА

$\hat{\theta}_n$ асимптотически нормальна, если $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d}_{n \rightarrow \infty} \eta$, где $\eta \sim N(0, \sigma^2(\theta))$ – нормальная случайная величина.
 σ – асимптотическая дисперсия.

ПРИМЕР: СИММЕТРИЧНЫЕ РАСПРЕДЕЛЕНИЯ

Пусть $p(x)$ – плотность (принимает неотрицательные значения, интегрируется в 1 по всей прямой); к тому же, $p(0) \neq 0$, $p(x)$ – чётная функция. Примеры: распределение нормальное (Гаусса) $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, распределение Коши $p(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, распределение Лапласа $p(x) = \frac{1}{2} e^{-|x|}$.

Соответствующая функция распределения $F(x) = \int_{-\infty}^x p(t) dt$ тогда симметричная.

Пусть функция распределения, породившая выборку, определена с точностью до сдвига $F_{\xi_i}(x) = F(x - \theta)$. Сдвиг можно найти по мат. ожиданию, либо по медиане распределения $ME = x_{1/2}$ (где $F(x_{1/2}) = \frac{1}{2}$). Более общее определение медианы – α -квантиль x_α . Тогда

$$\theta = E[\xi_i] = x_{1/2}.$$

Для медианы бы взяли оценку (конкурирующую) $\hat{\theta}_n^2 = ME = \begin{cases} x_{(k)}, & n = 2k - 1 \\ \frac{1}{2}x_{(k)} + \frac{1}{2}x_{(k+1)}, & n = 2k \end{cases}$

Возьмём оценку из мат. ожидания $\hat{\theta}_n^1 = E[\xi_i] = \bar{x}$.

Проверим на асимптотическую нормальность:

$$\left(\frac{\xi_1 + \dots + \xi_n}{n} - \theta \right) \sqrt{n} = \left(\frac{\xi_1 + \dots + \xi_n - nE[\xi_1]}{n} \right) \sqrt{n} = \frac{\xi_1 + \dots + \xi_n - nE[\xi_1]}{\sqrt{n}}$$

По центральной предельной теореме: $\frac{\xi_1 + \dots + \xi_n - nE[\xi_1]}{\sqrt{n} \sqrt{D[\xi_1]}} \xrightarrow{d}_{n \rightarrow \infty} \eta \sim N(0, 1)$

Тогда, для нашего случая, $\sigma^2 = D[\xi_1]$

ТЕОРЕМА

Если есть чётная плотность, элементы распределения имеют функцию распределения с неизвестным параметром сдвига $F_{\xi_i}(x) = F(x - \theta)$, теоретическая и выборочная медианы сходятся следующим образом:

$$\sqrt{n} (ME - x_{1/2}) \xrightarrow{d}_{n \rightarrow \infty} \eta \sim N\left(0, \frac{1}{4p^2(0)}\right)$$

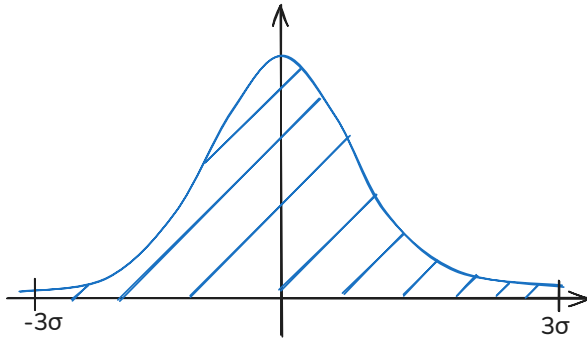
То есть $\sigma_{ME}^2 = \frac{1}{4p^2(0)}$.

Как выбирать оценку из асимптотически нормальных? Интуитивно понятно, что ту, у которой дисперсия меньше, или ту, которая быстрее сходится (требуется меньший объём выборки, чтобы аппроксимировать неизвестный параметр с заданной точностью).

УТВЕРЖДЕНИЕ

Чем меньше асимптотическая дисперсия, тем меньше необходимый объём выборки (при заданном ε)

Пусть есть плотность нормального распределения $N(0, \sigma^2) : p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$



Между $-3\sigma < x < 3\sigma$ вероятность $\sim 0,993$. Какой размер выборки n надо взять, чтобы $P(|\hat{\theta}_n - \theta| < \varepsilon) \approx 0.002$ для заданного ε ?

Если оценка асимптотически нормальная,

$$P\left(\sqrt{n}|\hat{\theta}_n - \theta| > \sqrt{n}\varepsilon\right) \rightarrow_{n \rightarrow \infty} 2 \int_{\varepsilon\sqrt{n}}^{\infty} \frac{e^{-\frac{x^2}{2\sigma^2(\theta)}}}{\sqrt{2\pi}\sigma^2(\theta)} dx$$

Видно, что надо взять $\varepsilon\sqrt{n} = 3\sigma(\theta)$, и

$$n = \frac{9\sigma^2(\theta)}{\varepsilon^2}.$$

≡ ПРИМЕР: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Пусть есть распределение $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Пусть функция распределения известна с точностью до сдвига $F_{\xi_i}(x) = F(x - \theta)$ (то есть $p_{\xi_i}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$)

Сравним дисперсии оценки по мат. ожиданию и по медиане.

$$\sigma_{\bar{x}}^2 = D\xi_1 = E[(\xi_1 - E[\xi_1])^2] = E[(\xi - \theta)^2] = \int_{-\infty}^{\infty} (x - \theta)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} dx$$

$$\text{Сделаем замену } x - \theta \rightarrow y : \sigma_{\bar{x}}^2 = \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1$$

$$\text{Теперь, по теореме, } \sigma_{\text{МЕ}}^2 = \frac{1}{4p^2(0)} = \frac{\pi}{2}.$$

Получается, $\bar{x}$ лучше, чем МЕ.

Лекция 4: Методы построения статистических оценок.

Если надо найти мат. ожидание величины ξ_i , было бы нормально оценить как среднее арифметическое: $E[\xi_i] = \bar{x} = \frac{1}{n} \sum_{i=1}^n \xi_i$.

Аналогично, для 2-го момента, можно воспользоваться средним квадратичным:

$$E[\xi_i^2] = \overline{x^2} = \frac{1}{n} \sum_{i=1}^n \xi_i^2.$$

Оценка момента

$$E[\xi_i^k] = \overline{x^k} = \frac{1}{n} \sum_{i=1}^n \xi_i^k$$

является несмещённой и состоятельной.

🔗 МЕТОД МОМЕНТОВ

В реальности распределение параметризовано m различными параметрами: $F_{\xi_i} \in \{F_{\theta}\}, \theta \in \Theta \subset \mathbb{R}^m$

Посчитанный k -ый момент – функция от параметров $E[\xi_i^k] = f_k(\theta_1, \dots, \theta_m)$

Составим систему уравнений:

$$\begin{cases} f_1(\theta_1, \dots, \theta_m) = \bar{x} \\ f_2(\theta_1, \dots, \theta_m) = \bar{x}^2 \\ \dots \\ f_m(\theta_1, \dots, \theta_m) = \bar{x}^m \end{cases}$$

Предположим, что отображение $\vec{f}(\vec{\theta})$ взаимно однозначно и взаимно непрерывно. Тогда решение системы $\hat{\theta}_1^n, \dots, \hat{\theta}_m^n$ единственно и существует.

Такое решение называется оценкой, полученной с помощью метода моментов. Она будет состоятельной.

≡ ПРИМЕР: БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$$\begin{aligned} \xi_i &\sim \text{Binom}(k, \theta) : P(\xi_i = a) = C_k^a \theta^a (1 - \theta)^{k-a} \\ E[\xi_i] &= k\theta =: f_1(\theta) \rightarrow k\theta = \bar{x} \rightarrow \hat{\theta}_n = \frac{1}{kn} \sum_{i=1}^n \xi_i \end{aligned}$$

≡ ПРИМЕР: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$$\begin{aligned} \xi_i &\sim N(\theta_1, \theta_2^2) \\ E[\xi_i] &= \bar{x} = \theta_1 =: f_1(\theta_1, \theta_2^2) \\ E[\xi_i^2] &= \bar{x}^2 = D[\xi_i] + E^2[\xi_i] = \theta_2^2 + \theta_1^2 =: f_2(\theta_1, \theta_2^2) \end{aligned}$$

Из системы следует, что $\begin{cases} \theta_1 = \bar{x} \\ \theta_1^2 + \theta_2^2 = \bar{x}^2 \end{cases} \rightarrow \begin{cases} \hat{\theta}_{n1} = \bar{x} \\ \hat{\theta}_{n2} = \sqrt{\bar{x}^2 - \bar{x}^2} \end{cases}$

≡ ПРИМЕР: РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ С НЕИЗВЕСТНЫМ ПРАВЫМ КОНЦОМ

$$\begin{aligned} \xi_i &\sim R(0, \theta) \text{ (плотность на отрезке } 1/\theta, \text{ вне отрезка } 0) \\ E[\xi_i] &= \frac{\theta}{2} \rightarrow \hat{\theta}_n = \frac{2}{n} \sum_{i=1}^n \xi_i \end{aligned}$$

🔗 МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Введём функцию

$$f(x; \vec{\theta}) := \begin{cases} P_\theta(\xi = x) & \text{дискретное распределение} \\ p_\xi(x; \vec{\theta}) & \text{абсолютно непрерывное распределение} \end{cases}$$

Для дискретного распределения это вероятность совпадения ξ, x при оценке θ (вероятность определяется функцией распределения, а она зависит от θ); для непрерывного – плотность.

Функция правдоподобия:

$$L(x_1, \dots, x_n; \vec{\theta}) = \prod_{i=1}^n f(x_i; \vec{\theta})$$

Оценка максимального правдоподобия (ОМП) – точка $\hat{\theta}_n$ максимума функции L .

Замечание: ОМП можно найти, приравняв все производные $\frac{\partial L}{\partial \theta_i}$ к нулю. Но дифференцировать обычно сложно из-за произведения функций. Можно прологарифмировать, и затем продифференцировать.

≡ ПРИМЕР: БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$$\xi_i \sim \text{Binom}(k, \theta) : P(\xi_i = a) = C_k^a \theta^a (1 - \theta)^{k-a}$$

Реализации случайных величин – числа от 0 до k . Тогда $P(\xi_i = x_i) = C_k^{x_i} \theta^{x_i} (1 - \theta)^{k-x_i}$

$$L(x_1, \dots, x_n; \vec{\theta}) = \prod_{i=1}^n C_k^{x_i} \theta^{x_i} (1 - \theta)^{k-x_i}$$

$$\ln L = \sum_{i=1}^n (\ln C_k^{x_i} + x_i \ln \theta + (k - x_i) \ln(1 - \theta))$$

$$\frac{\partial \ln L}{\partial \theta} = \sum_{i=1}^n \left(0 + \frac{x_i}{\theta} - \frac{k-x_i}{1-\theta} \right) = 0 \rightarrow \sum_{i=1}^n x_i - nk\theta = 0 \rightarrow \hat{\theta}_n = \frac{1}{kn} \sum_{i=1}^n x_i$$

Ответ такой же, как в методе моментов.

≡ ПРИМЕР: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$$\xi_i \sim N(\theta_1, \theta_2^2)$$

$$L(x_1, \dots, x_n; \theta_1, \theta_2^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2^2}}$$

$$\ln L = \sum_{i=1}^n \left(-\ln(\sqrt{2\pi}) - \ln \theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2^2} \right)$$

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = \sum_{i=1}^n \frac{(x_i - \theta_1)}{\theta_2^2} = 0 \\ \frac{\partial \ln L}{\partial \theta_2} = \sum_{i=1}^n \left(-\frac{1}{\theta_2} + \frac{(x_i - \theta_1)^2}{\theta_2^3} \right) = 0 \end{cases} \rightarrow \begin{cases} \hat{\theta}_1 = \bar{x} \\ \hat{\theta}_2^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$$

Из Лекции 1 помним, что такая оценка смещённая, но состоятельная.

Ответ такой же, как в методе моментов.

≡ ПРИМЕР: РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ С НЕИЗВЕСТНЫМ ПРАВЫМ КОНЦОМ

$$\xi_i \sim R(0, \theta)$$

$$f(x_i; \theta) = \frac{1}{\theta} I_{x_i \in [0, \theta]} = \frac{1}{\theta} I_{x_i \leq \theta} \text{ (левая граница будет, иначе задача бессмысленная)}$$

$$L(x_1, \dots, x_n; \theta_1, \theta_2^2) = \frac{1}{\theta^n} \prod_{i=1}^n I_{x_i \leq \theta}$$

Производной индикаторной функции нет, дифференцировать нельзя. Но из общих соображений ищем максимум: если хотя бы один $x_i > \theta$, то всё обратится в ноль. Получается, θ не меньше всех x_i (в терминах вариационного ряда $\theta \leq x_{(n)}$). Но чем больше значение θ , тем меньше коэффициент $\frac{1}{\theta^n}$.

Получается, $\hat{\theta}_n \leq x_{(n)}$.

Ответ отличается от метода моментов. Такая оценка смещённая.

Лекция 5: Условия регулярности. Доверительные интервалы.

Пусть есть 1 параметр: $F_{\xi_i} \in \{F_\theta\}$, $\theta \in \Theta \subset \mathbb{R}^1$

Вспомним обозначения:

$$f(x; \vec{\theta}) := \begin{cases} P_\theta(\xi = x) & \text{дискретное распределение} \\ p_\xi(x; \vec{\theta}) & \text{абсолютно непрерывное распределение} \end{cases}$$

Функция максимального правдоподобия: $L(x_1, \dots, x_n; \vec{\theta}) = \prod_{i=1}^n f(x_i; \vec{\theta})$

Введём множество $A := \{x : f(x; \theta) \neq 0\}$

Введём функцию $u = \frac{\partial \ln f(x; \theta)}{\partial \theta}$

Для нормального распределения $u = \frac{\partial \ln f(x; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left(-\ln \sqrt{2\pi} - \frac{(x - \theta)^2}{2} \right) = x - \theta$

🌀 УСЛОВИЯ РЕГУЛЯРНОСТИ

1. A не зависит от θ

Контрпример: $R(0, \theta)$

2. $f(x, \theta)$ дифференцируемо по θ на A

3.
$$\begin{cases} E[u] = 0 & \forall \theta \in \Theta, \\ 0 < D[u] < \infty. \end{cases}$$

Для непрерывного случая $E[u] = \int_A \frac{\partial \ln f(x; \theta)}{\partial \theta} f(x; \theta) dx = \int_A \frac{\partial f}{\partial \theta} dx$

Вспомним, что $\int_A f(x; \theta) dx = 1$. Очевидно тогда, $\frac{\partial}{\partial \theta} \int_A f(x; \theta) dx = 0$.

3-е условие регулярности говорит о том, что дифференцирование и интегрирование можно переставлять.

Из $E[u] = 0$ получаем $D[u] = E \left[\left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 \right] = I_1(\theta)$ (I_1 – информация, не индикатор)

ИНФОРМАЦИЯ ФИШЕРА

$$I_n(\theta) = E \left[\left(\frac{\partial \ln L(\xi_1, \dots, \xi_n; \theta)}{\partial \theta} \right)^2 \right]$$

Поскольку величины $\xi_1, \dots, \xi_n$ независимы, то сумма дисперсий равна дисперсии суммы, и $I_n(\theta) = nI_1(\theta)$

НЕРАВЕНСТВО РАО-КРАМЕРА

Пусть выполнены условия регулярности (1-3).

Пусть $\hat{\theta}_n$ – некоторая несмещённая оценка параметра θ .

Пусть $E[\hat{\theta}_n] = \int_{\mathbb{R}^n} \hat{\theta}_n(x_1, \dots, x_n) L(x_1, \dots, x_n; \vec{\theta}) dx_1 \dots dx_n$ можно дифференцировать по θ под знаком интеграла.

Тогда

$$D[\hat{\theta}_n] \geq \frac{1}{I_n(\theta)}.$$

Замечание: это значит, что нельзя придумать оценку со слишком быстро стремящейся к нулю дисперсией.

Замечание: $I_n(\theta) \sim n \rightarrow D[\hat{\theta}_n] \sim \frac{1}{n}$

Замечание: для равномерного распределения $R(0, \theta)$ для оценки максимального правдоподобия $x_{(n)}$ выполняется $D[\hat{\theta}_n] \sim \frac{1}{n^2}$. То есть оценка сверхэффективна.

ТЕОРЕМА

Пусть выполнены условия регулярности (1-3).

Пусть $f(x; \theta)$ трижды дифференцируема по θ , причём $f'''_{\theta\theta\theta}(\xi; \theta) \leq h(\xi)$, где $E[h(\xi)] < \infty$. (Райгородский: это формальность)

Тогда решение уравнения необходимого условия экстремума $\frac{\partial \ln f}{\partial \theta} = 0$ является оценкой максимального правдоподобия (не надо экстремум проверять на максимум), причём обладает эта оценка асимптотически нормальна с асимптотической дисперсией $\frac{1}{I_1(\theta)}$ и среди всех асимптотически нормальных оценок эта дисперсия глобально самая маленькая (это асимптотически эффективная оценка).

Мотивация интервальных оценок: у нас пока были только точечные. Теперь мы хотим приблизиться не одной функцией, а зажать между двумя функций. Получится доверительный интервал.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ

Доверительным интервалом с доверительной вероятностью $\alpha \in (0, 1)$ называется пара случайных величин $\hat{\theta}_n^1(\xi_1, \dots, \xi_n)$, $\hat{\theta}_n^2(\xi_1, \dots, \xi_n)$ такие, что вероятность того, что неизвестный параметр будет между ними, больше α :

$$\hat{\theta}_n^1, \hat{\theta}_n^2 : P(\theta \in [\hat{\theta}_n^1, \hat{\theta}_n^2]) \geq \alpha.$$

Асимптотические доверительные интервалы: $P(\theta \in [\hat{\theta}_n^1, \hat{\theta}_n^2]) \rightarrow_{n \rightarrow \infty} \alpha$.

Доверительные интервалы бывают односторонними.

ПРИМЕР: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$$\xi_i \sim N(\theta, \sigma^2)$$

Подгоним к другому распределению, где мы очень просто найдём квантили: $\frac{\xi_i - \theta}{\sigma} \sim N(0, 1)$ для одного элемента, $\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\xi_i - \theta}{\sigma} \right) \sim N(0, 1)$ для n элементов.

Квантиль стандартного нормального распределения u_α .

Возьмём такие γ, β , что $\gamma - \beta = \alpha$. По ним найдём u_γ, u_β , тогда $\alpha = F(u_\gamma) - F(u_\beta) = P(u_\beta \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\xi_i - \theta}{\sigma} \right) \leq u_\gamma)$

Выразим неравенство через θ :

$$\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{u_\gamma \sigma}{\sqrt{n}} \leq \theta \leq \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{u_\beta \sigma}{\sqrt{n}}$$

ПРИМЕР: НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$$\xi_i \sim N(a, \theta^2)$$

Рассмотрим распределение "хи-квадрат с N степенями свободы" $\chi_n^2 : \xi_1^2 + \dots + \xi_n^2$, где $\xi_i \sim N(0, 1)$.

Подгоняем под $\sum_{i=1}^n \left(\frac{\xi_i - a}{\theta} \right)^2 \sim \chi_n^2$.

Находим квантили $\chi_{n,\gamma}^2, \chi_{n,\beta}^2$, далее аналогично предыдущему примеру.