Конспект по курсу

Математическая статистика

Contributors:
Андрей Степанов
Алексей Журавлев

Лектор:
Шабанов

МФТИ

Последнее обновление: 21 апреля 2015 г.

Содержание

1 Введение. Сходимости векторов.

1.1 Введение.

Математическая статистика – это раздел теории вероятностей, который решает обратные задачи к классическим задачам в теории вероятностей.

Типичная задача в теории вероятностей – это найти или оценить характеристики случайного эксперимента, зная его природу случайности.

Типичная задача в математической статистике – по данным результатов случайного эксперимента выяснить природу его случайности.

Пример (Классический пример)**.** В городе есть n жителей, m из которых болеют. Считаем, что n дано заранее.

- Задача ТВ: с какой вероятностью при известном m в случайной выборке из a жителей будет b заболевших

- Задача МС: известно, что в выборке из a жителей оказалось b заболевших. Как в этом случае можно оценить m

Пример (Выборка)**.** Предположим, что мы проводим эксперимент. Пусть дан какой-то физический прибор, и пусть ξ – случайная величина, описывающая результат измерения этим прибором, $\xi \sim P_\xi$ (ξ имеет распределение P_ξ). Например, если прибор – это счетчик Гейгера, то ξ – это уровень радиации, им зарегестрированный. Давайте также считать, что на время эксперимента распределение ξ не меняется, и результат измерения прибора не зависит от предыдущих измерений. Пусть $X_1, \ldots, X_n$ – эти результаты измерения в какие-то моменты времени. На языке теории вероятностей это можно переформулировать так: $X_1, \ldots, X_n$ – *реализации независимых одинаково распределенных случайных величин* $\xi_1, \ldots, \xi_n$.

Задача состоит в том, чтобы оценить $E\xi$ по этим самым $X_1, \ldots, X_n$.

Пример (Регрессионная модель)**.** Пусть материальная точка движется по прямой, стартовав из точки x_0 с постоянной скоростью, равной v_0. Мы их не знаем, и будем считать, что это случайные величины. Пусть $x_1, \ldots, x_n$ – это измеренные нами положения этой материальной точки в моменты времени $t_1, \ldots, t_n$ соответственно. Или, по другому, $x_1, \ldots, x_n$ – это реализации случайных величин $\xi_1, \ldots, \xi_n$, причем ξ_i отвечает за измеренный нами результат положения точки в момент времени t_i. Понятно, что эти случайные величины уже будут зависимы. Дополнительно положим, что *погрешность измерения подчиняется нормальному распределению.* То есть: $\xi_i = x_0 + v_0 \cdot t_i + \varepsilon_i$, где ε_i – нормально распределенная случайная величина, отвечающая за ошибку i-того измерения.

Задача заключается в том, чтобы оценить x_0 и v_0 по этим данным $(x_0, \ldots, x_n, t_0, \ldots, t_n)$

Пример (Проверка на однородность)**.** Пусть $X_1, \ldots, X_n$ – это результаты эксперимента в условиях A, а $Y_1, \ldots, Y_m$ – результаты того-же самого эксперимента в условиях B. Нужно выяснить, влияют ли эти условия на результат. (Если для сокращения записи отождествить результат эксперимента со случайной величиной, реализацией которой он является, а также считать, что $X_i \sim X$, $Y_i \sim Y$ то можно записать так: $X \overset{d}{\sim} Y$?)

Замечание. Как мы видим, задача матстатистики – представить оптимальное решение на основе статистических данных. Типичная характерная черта задач – это довольно большое количество дополнительных ограничений на природу явлений (независимость и одинаковая распределенность результата, нормальное распределение погрешностей и т.д.). Такие ограничения в реальных условиях иногда бывает трудно проверить, поэтому нужно быть крайне внимательным при применении какого-либо результата из матстатистики в реальных задачах. Однако такое требование к внимательности компенсируется тем, что результаты из матстатистики находят широкое применение в экспериментальной физике, машинном обучении, data mining и прочих облостях науки.

1.2 Сходимости случайных векторов

Определение 1.1. Пусть $\{\xi_n : n \in \mathbb{N}\}$ – последовательность случайных векторов из $\mathbb{R}^m$. ξ – случайный вектор из $\mathbb{R}^m$. Говорят что:

- ξ_n сходится к ξ почти наверное (обозначение: $\xi_n \overset{\text{п.н.}}{\to} \xi$), если :

$$P(\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\}) = 1$$

- ξ_n сходится к ξ по вероятности (обозначение $\xi_n \overset{P}{\to} \xi$), если:

$$\forall\varepsilon > 0 : \lim_{n\to\infty} P(\{\omega : \|\xi_n(\omega) - \xi(\omega)\| > \varepsilon\}) = 0$$

- ξ_n сходитіся к ξ по распределению (обозначение $\xi_n \overset{d}{\to} \xi$), если:

  $\forall f$ – ограниченной непрерывной функции $\mathbb{R}^m \mapsto \mathbb{R} : \lim\limits_{n\to\infty} Ef(\xi_n) = Ef(\xi)$

Утверждение 1.1. *Пусть* $\xi_n = (\xi_n^1, \dots, \xi_n^m), \xi = (\xi^1, \dots, \xi^m)$. *Тогда:*

- $\xi_n \overset{\text{п.н.}}{\to} \xi \Leftrightarrow \forall j : \xi_n^j \overset{\text{п.н.}}{\to} \xi^j$
- $\xi_n \overset{P}{\to} \xi \Leftrightarrow \forall j : \xi_n^j \overset{P}{\to} \xi^j$

Замечание. Для сходимости по распределению это утверждение не верно

Определение 1.2. Функции распределения F_{ξ_n} называются слабо сходящимися к F_ξ (обозначение: $F_{\xi_n} \overset{w}{\to} F_\xi$), если:

$\forall f$ – ограниченной непрерывной функции $\mathbb{R} \mapsto \mathbb{R} : \int_\Omega f(x)dF_{\xi_n} \to \int_\Omega f(x)dF_\xi$

Теорема 1.2 (Александрова). *Пусть* $\xi, \xi_1, \xi_2, \dots$ *– случайные величины. Тогда следующие утверждения эквивалентны:*

1. $\xi_n \xrightarrow{d} \xi$
2. $F_{\xi_n} \xrightarrow{w} F_\xi$
3. $\forall x$ *– точка непрерывности* $F_\xi : \lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$

Теорема 1.3 (многомерный случай, более слабая). *Пусть* $\xi, \xi_1, \xi_2, \dots$ *– случайные векторы из* $\mathbb{R}^m$*. Пусть* F_ξ *непрерывна. Тогда* $\xi_n \xrightarrow{d} \xi \Leftrightarrow \forall x \in \mathbb{R}^m : \lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$

1.3 Предельные теоремы

Теорема 1.4 (Закон больших чисел). *Пусть* $\xi_1, \xi_2, \dots$ *– попарно-некоррелированные одинаково распределенные случайные величины,* $D\xi_i$ *конечна,* $S_n = \xi_1 + \dots + \xi_n$*. Тогда:*

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0$$

, причем $ES_n = n \cdot a$*,* $a = E\xi_i$*.*

Теорема 1.5 (Усиленный закон больших чисел). *Пусть* $\xi_1, \xi_2, \dots$ *– независимые одинаково распределные случайные величины (или векторы),* $E\xi_i = a$ *– конечно,* $S_n = \xi_1 + \dots + \xi_n$*. Тогда:*

$$\frac{S_n}{n} \xrightarrow{\text{п.н.}} a$$

Теорема 1.6 (Центральная предельная теорема). *Пусть* $\xi_1, \xi_2, \dots$ *– независимые одинаково распределные случайные величины,* $E\xi_i = a$ *– конечно,* $0 < D\xi_i = \sigma^2$ *– тоже конечно,* $S_n = \xi_1 + \dots + \xi_n$*. Тогда:*

$$\sqrt{n}\left(\frac{S_n}{n} - a\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

Теорема 1.7 (Многомерная центральная предельная теорема). *Пусть* $\xi_1, \xi_2, \dots$ *– независимые одинаково распределные случайные векторы,* $E\xi_i = a$ *– конечно,* $0 < D\xi_i = \Sigma$ *– матрица ковариаций, тоже конечна,* $S_n = \xi_1 + \dots + \xi_n$*. Тогда:*

$$\sqrt{n}\left(\frac{S_n}{n} - a\right) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

Утверждение 1.8. *Пусть* $\xi, \xi_1, \xi_2, \dots$ *– случайные векторы. Тогда:*

1. $\xi_n \xrightarrow{\text{п.н.}} \xi \Rightarrow \xi_n \xrightarrow{P} \xi$
2. $\xi_n \xrightarrow{P} \xi \Rightarrow \xi_n \xrightarrow{d} \xi$

Доказательство.

1. Следует из того, что векторная сходимость эквивалентна покоординатной, а для координат верны одномерные аналоги.

2. Доказывается аналогично одномерному случаю.

□

Лемма 1.9 (о сходящейся подпоследовательности). *Пусть* $\xi_n \xrightarrow{P} \xi$. *Тогда* $\exists$ *подпоследовательность* $\xi_{n_k} : \xi_{n_k} \xrightarrow{п.н.} \xi$

Теорема 1.10 (о наследовании сходимостей). *Пусть* $\xi, \xi_1, \xi_2, \ldots$ *– случайные векторы из* $\mathbb{R}^m$. *Пусть также* $h : \mathbb{R}^m \mapsto \mathbb{R}^s$ *– функция, непрерывная почти всюду относительно распределения* ξ *(то есть* $\exists B \in \mathcal{B}(\mathbb{R}^m), P_\xi(B) = 1 : h$ *непрерывна на* B. *Тогда:*

1. $\xi_n \xrightarrow{п.н.} \xi \Rightarrow h(\xi_n) \xrightarrow{п.н.} h(\xi)$

2. $\xi_n \xrightarrow{P} \xi \Rightarrow h(\xi_n) \xrightarrow{P} h(\xi)$

3. *Если дополнительно* h *непрерывна всюду, а не почти всюду, то:* $\xi_n \xrightarrow{d} \xi \Rightarrow h(\xi_n) \xrightarrow{d} h(\xi)$

Доказательство.

1. $1 = P(\{\omega : \xi_n(\omega) \to \xi(\omega)\}) \le P(\{\omega : h(\xi_n)(\omega) \to h(\xi)(\omega)\})$, так как $\{\omega : \xi_n(\omega) \to \xi(\omega)\} \subset \{\omega : h(\xi_n)(\omega) \to h(\xi)(\omega)\}$

2. Пусть не выполнено, что $h(\xi_n) \xrightarrow{P} h(\xi)$. Тогда $\exists \varepsilon, \delta$ а так же подпоследовательность ξ_{n_k}:

$$\forall k : P(\|h(\xi_{n_k}) - h(\xi)\| \ge \varepsilon) \ge \delta$$

. Так как $\xi_{n_k} \xrightarrow{P} \xi$, то существует подпоследовательность $\xi_{n_{k_l}} : \xi_{n_{k_l}} \xrightarrow{\text{п.н.}} \xi$. Тогда согласно первому пункту $h(\xi_{n_{k_l}}) \xrightarrow{\text{п.н.}} h(\xi)$. Но такого быть не может, т.к. $\forall k : P(\|h(\xi_{n_k}) - h(\xi)\| \ge \varepsilon) \ge \delta$

3. Пусть $f : \mathbb{R}^m \mapsto \mathbb{R}$ – непрерывная ограниченная функция. Тогда $g = f \circ h$ – тоже непрерывная ограниченная функция. Так как $\xi_n \xrightarrow{d} \xi$, то $Eg(\xi_n) \to Eg(\xi)$. А значит, $Ef(h(\xi_n)) \to Ef(h(\xi))$

□

Лемма 1.11 (Слуцкого). *Пусть* $\xi_n \xrightarrow{d} \xi$ *– случайные величины, а* $\eta_n \xrightarrow{d} C = const$. *Тогда* $\xi_n + \eta_n \xrightarrow{d} \xi + C$ *и* $\xi_n \cdot \eta_n \xrightarrow{d} \xi \cdot C$.

Доказательство. Пусть t – точка непрерывности функции распределения $F_{\xi+C}$ случайной величины $\xi + C$. Докажем только для суммы, для произведения аналогично. Пусть $\varepsilon > 0$ такое,что $t \pm \varepsilon$ – тоже точки непрерывности

$F_{\xi+C}$. Мы хотим показать, что $F_{\xi_n+\eta_n}(t) \to F_{\xi+C}(t)$. Будем для этого пользоваться тем, что $\eta_n \xrightarrow{d} C \Leftrightarrow \eta_n \xrightarrow{P} C$.

$$P(\xi_n + \eta_n \leq t) = P(\xi_n + \eta_n \leq t, \eta_n \geq C - \varepsilon) + P(\xi_n + \eta_n \leq t, \eta_n < C - \varepsilon) \leq \\ P(\xi_n \leq t - C + \varepsilon) + P(\|\eta_n - C\| > \varepsilon) = F_{\xi_n+c}(t+\varepsilon) + P(\|\eta_n - C\| > \varepsilon)$$

Но $\xi_n \xrightarrow{d} \xi \Rightarrow \xi_n + C \xrightarrow{d} \xi + C$. Кроме того, $t+\varepsilon$ – точка непрерывности $F_{\xi+C}$

$$\limsup_{n\to\infty} F_{\xi_n+\eta_n}(t) \leq \lim_{n\to\infty} F_{\xi_n+c}(t+\varepsilon) + \lim_{n\to\infty} P(\|\eta_n - C\| > \varepsilon) = F_{\xi+C}(t+\varepsilon)$$

Аналогично,

$$1 - f_{\xi_n+\eta_n}(t) \leq P(\|\eta_n - C\| > \varepsilon) + 1 - F_{\xi_n+C}(t-\varepsilon)$$

Откуда

$$\liminf_{n\to\infty} F_{\xi_n+\eta_n}(t) \geq F_{\xi+C}(t-\varepsilon)$$

В силу произвольности $\varepsilon > 0$ и того факта, что t – точка непрерывности функции распределения $F_{\xi+c}$ получаем, что

$$\lim_n F_{\xi_n+\eta_n} = F_{\xi+C}(t)$$

□

Утверждение 1.12 (применение леммы Слуцкого)**.** *Пусть $\xi_n \xrightarrow{d} \xi$ – случайные величины. $h(x)$ – дифференцируема в точке $a \in \mathbb{R}$. $\{b_n > 0\}$ – числовая последовательность, $b_n \to 0$. Тогда:*

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} h'(a)\xi$$

Доказательство. По лемме Слуцкого $\xi_n b_n \xrightarrow{d} \xi_n \cdot 0 = 0$. Рассмотрим

$$H(x) = \begin{cases} \frac{h(x+a)-h(a)}{x}, x \neq 0 \\ h'(a), x = 0 \end{cases}$$

$H(x)$ непрерывна в 0 по определению, а также непрерывна на $\mathbb{R} \setminus 0$ как композиция непрерывных функций. По теореме о наследовании сходимостей $H(\xi_n b_n) \xrightarrow{d} H(0) = h'(a)$. По лемем слуцкого $\xi_n H(\xi_n b_n) \xrightarrow{d} h'(a)\xi$ □

Теорема 1.13 (многомерный вариант)**.** *Пусть $\xi_n \xrightarrow{d} \xi$ – случайные векторы. $h(x)$ – дифференцируема в точке $a \in \mathbb{R}$ (то есть существует матрица частных производных, или матрица Якоби $J(h)$). $\{b_n > 0\}$ – числовая последовательность, $b_n \to 0$. Тогда:*

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} J(h)(a)\xi$$

2 Вероятностно-статистические модели и выборки

2.1 Вероятностно-статистическая модель

Определение 2.1. Множество всех возможных значений наблюдения называется выборочным пространством и обозначается $\mathfrak{X}$

Определение 2.2. Наблюдение X – это результат случайного выбора элемента из выборочного пространства. Наша цель – по наблюдению X сделать выводы о его распределении P.

Определение 2.3. Если $X = (X_1, \dots, X_n)$ – набор независимых одинаково-распределенных случайных величин имеющих распределению P, то X называется выборкой размера n из неизвестного распределения P

Замечание. В дальнейшем будем обозначать: $X = (X_1, \dots, X_n)$ – выборка на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ с неизвестным распределением P на $(\mathfrak{X}, \mathcal{B}_x)$ (например, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Определение 2.4. Для каждого множества $B \in \mathcal{B}_x$) введем $P_n^*(B) = \frac{\nu_n(B)}{n}$, где $\nu_n(B)$ – это количество элементов из $X_1, \dots, X_n$, которые попали в B. То есть формально:

$$P_n^*(B) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \in B\}$$

. Такая величина называется эмпирическим распределением.

Утверждение 2.1. *Пусть P – неизвестное распределение X_i. Тогда $\forall B \in \mathcal{B}_x : P_n^*(B) \overset{п.н.}{\to} P(B)$.*

Доказательство. Фиксируем $B \in \mathcal{B}_x$. Тогда для фиксированного B индикаторы $I\{X_i \in B\}$ будут являтся случайными величинами, причем независимыми и одинаково распределенными, поскольку исходные случайные величины были независимыми и одинаково распределенными. Введем

$$S_n = \sum_{i=1}^{n} I\{X_i \in B\}$$

По усиленному закону больших чисел:

$$\frac{1}{n}S_n \overset{\text{п.н.}}{\to} P(X_1 \in B) = P(B)$$

□

Определение 2.5. Пусть $X = (X_1, \dots, X_n)$ – выборка.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \le x\}$$

называется эмпирической функцией распределения (она является функцией распределения для эмпирического распределения P_n^*).

Замечание.

$$\forall x \in \mathbb{R} : F_n(x) \overset{\text{п.н.}}{\to} F(x)$$

Теорема 2.2 (Гливенко - Кантелли)**.** *Если $F(x)$ – функция распределения элементов выборки $X_1, \ldots, X_n$ (то есть функция распределения для распределения P), то:*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{п.н.}{\to} 0$$

Доказательство. Зафиксируем элементарный исход $\omega \in \Omega$. Тогда случайные величины $X_1, \ldots, X_n$ превращаются в числа. Посмотрим на функцию распределения $F_n(x)$. Она является непрерывной справа, так как является конечной суммой индикаторов вида $I\{X_i \leq x\}$, а $F(x)$ непрерывна справа как функция распределения. Модуль их разности тоже непрерывен справа. Тогда $\sup_{x\in\mathbb{R}} |F_n(x)-F(x)| = \sup_{x\in\mathbb{Q}} |F_n(x)-F(x)|$ – это супремум счетного числа случайных величин. Поэтому $|F_n(x)-F(x)|$ тоже является случайной величиной.

Пусть $N \in \mathbb{N}$ – достаточно большое. Для каждого $k \in \{1, \ldots, N-1\}$ введем $x_{k,N} = \inf\{x : F(x) \geq \frac{k}{N}\}$. Полагаем также $x_{0,N} = -\infty, x_{N,N} = +\infty$. Оценим $F_n(x) - F(x)$ для $x \in [x_{k,N}, x_{k+1,N})$:

$$\begin{aligned}
F_n(x) - F(x) &\leq F_n(x_{k+1,N} - 0) - F(x_{k,N}) = \\
&= F_n(x_{k+1,N} - 0) - F(x_{k+1,N} - 0) + F(x_{k+1,N} - 0) - F(x_{k,N}) \leq \\
&\qquad F_n(x_{k+1,N} - 0) - F(x_{k+1,N} - 0) + \frac{1}{n}
\end{aligned}$$

Совершенно аналогично показывается, что $F_n(x)-F(x) \geq F_n(x_{k,N})-F(x_{k,N})-\frac{1}{N}$ Откуда получаем, что:

$$\sup_{x\in\mathbb{Q}} |F_n(x)-F(x)| \leq \max_{k,l}\{|F_n(x_{k,N}-0)-F(x_{k,N}-0)|, |F_n(x_{l,N})-F(x_{l,N})|\}+\frac{1}{N}$$

Зафиксируем $\varepsilon > 0$ и возьмем $N : \frac{1}{N} < \varepsilon$. По усиленному закону больших чисел $F_n(x) \overset{\text{п.н.}}{\to} F_(x), F_n(x-0) \overset{\text{п.н.}}{\to} F(x-0)$. То есть $\forall x \in \mathbb{Q} : P(\limsup_n |F_n(x)-F(x)| > \varepsilon) = 0$. А значит, $P(\sup_{x\in\mathbb{Q}} \limsup_n |F_n(x)-F(x)| > \varepsilon) = 0$. Поменяв местами sup и lim sup и считая $\varepsilon = \frac{1}{m}$, получаем:

$$\forall m : P(\limsup_n \sup_{x\in\mathbb{Q}} |F_n(x) - F(x)| > \frac{1}{m}) = 0$$

Пользуясь теоремой о непрерывности вероятностной меры, имеем:

$$P(\lim_n \sup_{x\in\mathbb{Q}} |F_n(x) - F(x)| > 0) = 0$$

$\square$

Замечание. Пусть на $\mathfrak{X}$ задана σ-алгебра $\mathcal{B}_x$. Как правило $(\mathfrak{X}, \mathcal{B}_x) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ Наблюдение X – это, формально, тождественная случайная величина на $(\mathfrak{X}, \mathcal{B}_x, P)$. Обычно про распределение $P = P_X$ известно, что оно принадлежит некому классу распределений $\mathcal{P}$, например, классу нормальных распределений

Определение 2.6. Тройка $(\mathfrak{X}, \mathcal{B}_x, \mathcal{P})$, где $\mathcal{P}$ – это класс вероятностных мер на $(\mathfrak{X}, \mathcal{B}_x)$, называется вероятностно-статистической моделью

Замечание. $\forall P \in \mathcal{P} : (\mathfrak{X}, \mathcal{B}_x, P)$ – вероятностное пространство

Определение 2.7. Вероятностно-статистическая модель $(\mathfrak{X}, \mathcal{B}_x, \mathcal{P})$ называется параметрической, если класс $\mathcal{P}$ параметризован, то есть $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Также считаем, что $P_{\theta_1} \neq P_{\theta_2}$, если $\theta_1 \neq \theta_2$ asdasdasd

2.2 Моделирование выборки

2.2.1 Конечная выборка

Мы хотим смоделировать конечную выборку $(X_1, \ldots, X_n)$ в терминах веротяностно-статистической модели. Пусть X_i принимает значения из $\mathfrak{X}$ и имеет неизвестное распределение $P \in \mathcal{P}$. В этом случае удобно рассмотреть следующую статистическую модель: $(\mathfrak{X}^n, \mathcal{B}_x^n, \mathcal{P}^n)$, где $\mathfrak{X}^n = \mathfrak{X} \times \ldots \times \mathfrak{X}$, $\mathcal{B}_x^n = \sigma(B_1 \times \ldots \times B_n : \forall i : B_i \in \mathcal{B}_x)$, $\mathcal{P}^n = \{P^n : P \in \mathcal{P}\}$, $P^n(B_1 \times \ldots \times B_n) = P(B_1) \cdot \ldots \cdot P(B_n)$. То есть в качестве выборочного пространства мы берем декартову степень, в качестве сигма алгебры, как и в случае с $\mathcal{B}(\mathbb{R}^n)$, минимальную сигма-алгебру, порожденную декартовыми произведениями всех измеримых множеств, а в качестве класса мер – класс степеней всех мер, где степень меры означает естественное продолжение одномерной меры на многомерное пространство. Какими же выбрать $(X_1, \ldots, X_n)$? Это просто: $X_i : \mathfrak{X}^n \mapsto \mathfrak{X}$ такое что $X_i((x_1, \ldots, x_n)) = x_i$.

2.2.2 Счетная выборка

TO BE CONTINUED...

Замечание. Мы будем опускать индексы у $\mathfrak{X}$, $\mathcal{B}_x$, $\mathcal{P}$ в целях удобства

Замечание. В параметрической модели вопрос о неизвестном распределении P_θ сводится в вопросу о значении $\theta \in \Theta$

2.3 Статистики и оценки

Пусть X – наблюдение со значениями из $\mathfrak{X}$ и неизвестным распределением P_θ, где $\theta \in \Theta$

Определение 2.8. Статистикой $S(X)$ называется измеримая функция от результатов наблюдения, то есть: $S : \mathfrak{X} \mapsto E$, где $(E, \mathcal{E})$ – измеримое пространство, а S является $\mathcal{E}|\mathcal{B}_x$ измеримой.

Если $S : \mathfrak{X} \mapsto \Theta$, то S называется оценкой параметра из Θ.

Пример.

1. Если $g : \mathbb{R} \mapsto \mathbb{R}$ – борелевская функция, то $\overline{g(x)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$ называется выборочной характеристикой функции $g(x)$ (среднее значение по элементам выборки)

3 Оценки и их свойства

Определение 3.1. $\overline{X^k}$ – – выборочный k-тый момент.

Определение 3.2. $S^2 = \overline{X^2} - \overline{X}^2$ – выборочная дисперсия.
Выборочный k-тый центральный момент $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$.
$X_(k)$ – выборочная k-тая порядковая статистика.

Определение 3.3. Квантиль z_p уровня $p \in (0, 1)$ функции распределения F – $\min\{x : F(x) > p\}$

Определение 3.4. Выборочная квантиль $\hat{z}_p$ – это квантиль эмпирической функции распределения.

Определение 3.5. Медиана распределения μ – это квантиль уровня $1/2$

Определение 3.6. Выборочная медиана $\overline{\mu}$ – это

$$\begin{cases} X_{(n/2)}, \text{ если } n \text{ – четно} \\ \frac{X_{(\lfloor n/2 \rfloor)} + X_{(\lceil n/2 \rceil)}}{2}, \text{ иначе} \end{cases}$$

3.1 Свойства оценок

Определение 3.7. Оценка $\hat{\theta}$ называется несмещенной, если $\forall \theta \in \Theta : E_\theta \hat{\theta} = \theta$

Определение 3.8. Оценка $\hat{\theta}_n = \theta_n(X_1, \cdots, X_n)$ называется состоятельной, если $\forall \theta \in \Theta : \hat{\theta}_n \xrightarrow{P} \theta$

Определение 3.9. Оценка $\hat{\theta}_n = \theta_n(X_1, \cdots, X_n)$ называется сильно состоятельной, если $\forall \theta \in \Theta : \hat{\theta}_n \xrightarrow{\text{п.н.}} \theta$

Определение 3.10. Оценка $\hat{\theta}_n = \theta_n(X_1, \cdots, X_n)$ называется асимптотически нормальной оценкой параметра θ, если $\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

Пример.

1. $\overline{X}$ – несмещенная оценка параметра θ семейства распределений $\mathcal{N}(\theta, \sigma^2)$

2. Более того, по УЗБЧ $\overline{X}$ – сильно состоятельная оценка θ

3.2 Методя нахождения оценок

3.2.1 Метод подстановки

$\theta = F(P_\theta)$ Например, если $\{P_\theta\} = \{U[0,\theta], \theta > 0\}$, тогда $P_\theta([0,1]) = \frac{1}{\theta}$ и $\theta = \frac{1}{P_\theta([0,1])}$ Тогда используя метод подстановки (подставляя эмпирическое распределение, вместо неизвестного распределения P_θ) получаем оценку $\hat{\theta} = \frac{1}{P^*(\theta)}$

3.2.2 Метод моментов

Утверждение 3.1. *Если m^{-1} непрерывна – то $\hat{\theta}_n$ – сильно состоятельная оценка*

Доказательство. По УЗБЧ $\overline{g}_i(X) \overset{\text{п.н.}}{\to} m_i(\theta)$. Так как m^{-1} непрерывная, то по теореме о наследовании $\hat{\theta}_n = m^{-1}(\overline{g_1(X)}, \cdots, \overline{g_k(X)})$ □

Утверждение 3.2. *Аналогично, $\hat{\theta}_n$ является асимптотически нормальной оценкой.*

Пример. $X_1, \cdots, X_n \sim \Gamma(\alpha, \lambda)$

Замечание. Метод моментов – это частный случай метода подстановки.

Теорема 3.3 (теорема об асимптотической нормальности выборочной квантили). *Пусть $X_1, \cdots, X_n \sim P$ с плотностью $f(x)$, пусть также $f(x)$ – непрерывно дифференцируема в некоторой окрестности z_p, где z_p – это квантиль уровня p распределения P. Пусть $f(x) > 0$ для всех $x \in \mathbb{R}$. Тогда $\sqrt{n}(\hat{z}_p - z_p) \xrightarrow{d} \mathcal{N}(0, \frac{p(1-p)}{(f(z_p))^2})$*

Пример. По теореме о асимптотической нормальности выборочной медианы $\hat{\mu}$ – а.н. оценка параметра θ распределения с плотностью $f(x) = \frac{1}{\pi(1+(x-\theta)^2)}$ с выборочной диспресией $\frac{\pi^2}{4}$.

Теорема 3.4. *Если τ – непрерывная функция на Θ, $\hat{\theta}_n$ – (сильно) состоятельная оценка параметра θ, то $\tau(\hat{\theta}_n)$ – сильно состоятельная оценка параметра $\tau(\theta)$*

Теорема 3.5. *Если $\hat{\theta}_n$ – асимптотически нормальная оценка параметра θ, τ – дифференцируема на Θ, то $\tau(\hat{\theta}_n)$ – асимптотически нормальная оценка параметра $\tau(\theta)$ с асимптотической дисперсией $\sigma^2(\theta)[\tau'(\theta)]^2$, где $\sigma^2(\theta)$ – асимптотическая дисперсия $\hat{\theta}_n$*

Доказательство. Используем теорему из первой лекции $h = \tau, b_n = frac1\sqrt{n}, a = \theta, \xi$ □

4 Сравнение оценок

Определение 4.1. Пусть X – наблюдение с неизвестным распределением $P \in \{P_\theta : \theta \in \Theta\}$. $\rho(x, y)$ – функция потерь. Тогда функцией риска оценкаи $\hat{\theta}(X)$ неизвестного параметра θ называется: $R(\hat{\theta}(X), \theta) = E_\theta \rho(\hat{\theta}(X), \theta)$

4.1 Равномерный подход

Определение 4.2. Оценка $\hat{\theta}(X)$ называется лучшей оценки $\theta^*(X)$ в равномерном подходе, если $\forall \theta \in \Theta : R(\hat{\theta}(X), \theta) \le R(\theta^*(X), \theta)$ и для некоторого $\theta \in \Theta$ неравенство строгое.

Определение 4.3. Если оценка $\hat{\theta}(X)$ лучше любой другой оценки в каком-либо классе оценок, то она называется наилучшей в этом классе

Замечание. Равномерный подход с квадративной функцией потерь называется среднеквадратическим подходом. Не для любого класса можно отыскать наилучшую оценку.

Определение 4.4. K – несмещенные оценки $\tau(\theta)$. В таком классе K со среднеквадратичной функцией потерь:

$$R(\hat{\theta}(X), \theta) = E_\theta(\hat{\theta}(X) - \theta)^2 = E_\theta(\hat{\theta}(X) - \tau(Theta))^2 + (\tau(\theta) - \theta)^2 = D_\theta \hat{\theta}(X)$$

Определение 4.5. Оценка называется допустимой, если для неё не существует лучшей оценки в равномерном подходе.

4.2 Байесовский подход

Пусть $R(\hat{\theta}, \theta)$ – функция риска для оценки $\hat{\theta}$, и задано Q – нек. распределение вероятностей на Θ. Тогда можно определить $R(\hat{\theta}) = \int_\Theta R(\hat{\theta}, t) Q(dt)$. Если Q имеет плотность $q(t)$, то $R(\hat{\theta}) = \int_\Theta R(\hat{\theta}, t) q(t) dt$

Определение 4.6. Если $R(\hat{\theta}) = \min_{\theta^* \in K} R(\theta^*)$, то $\hat{\theta}$ называется наилучшей в байесовском подходе в классе K.

Байесовские оценки являются допустимыми.

4.3 Минимаксный подход

Если $\hat{R}(\hat{\theta}) = \min_{\theta^* \in K} \sup_{\theta \in \Theta} R(\theta^*, \theta)$, то $\hat{\theta}$ называется наилучшей в минимаксном подходе в классе K.

4.4 Асимптотический подход

$X = (X_1, \cdots, X_n)$ – выборка растущего размера.

Пусть $\hat{\theta}_1$, $\hat{\theta}_2$ – две асимптотические оценки $\theta \in \Theta \subset \mathbb{R}$. Пусть $\sigma_1^2(\theta)$, $\sigma_2^2(\theta)$ – их асимптотические оценки. Мы будем говорить, что $\hat{\theta}_1$ лучше $\hat{\theta}_2$, если $\forall \theta \in \Theta : \sigma_1^2(\theta) \le \sigma_2^2(\theta)$ и для некоторого $\theta \in \Theta$ неравенство строое.

Оценка называется наилучшей в асимптотическом подходе в каком-то классе, если она лучше любой другой оценки в каком-то классе.

Пример. $X = (X_1, \cdots, X_n)$ – выборка из номального распределения с параметрами $(\theta, 1)$. Нужно сравинить в асимптотическом подходе оценки $\overline{X}$ и $\hat{\mu}$.

$\sqrt{n}(\overline{X} - \theta) \sim N(0,1)$ по ЦПТ.

По теореме об асимптотической нормальности выборочного квантиля: $\sqrt{n}(\hat{\mu} - \theta) \sim N(0, \frac{1}{1/4 \cdot p(\theta)^2} = N(0, \frac{\pi}{2})$.

Получили, что $\overline{X}$ лучше $\hat{\mu}$

4.5 Понятие плотности дискретного распределения

Функция $p(x) \geq 0$ называется плотностью вероятностной меры P на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, если $\forall B \in \mathcal{B}(\mathbb{R}) : P(B) =\in_B p(x)dx$. В этом случае P называется абсолютно непрерывной вероятностной мерой, $p(x)$ – плотность по мере Лебега.

Определение 4.7. Функция $\mu : \mathcal{B}(\mathbb{R}) \mapsto \mathbb{Z} \cup \{\infty\}$ определенная по правилу:

$$\mu(B) = \sum_{k \in \mathbb{Z}} I\{k \in B\}$$

называется считающей мерой на $\mathbb{Z}$.

Определение 4.8. Интегралом от функции $f(x)$ по считающей мере μ называется $\int_{\mathbb{R}} f(x)\mu(dx) := \sum_{k \in \mathbb{Z}} f(k)$

Для такого интеграла выполнены все основные свойства: линейность, сохранение отношения порядка, теоремы о сходимости и так далее. Аналогично можно определить считающую меру в $\mathbb{Z}^n$ и интеграл по ней.

Определение 4.9. Пусть ξ – случайная величина со значениями в $\mathbb{Z}$. Тогда её плотность по считающей мере μ называется $p(x) = P(\xi = x)$

Следствие. *Для любой функции $g(x)$ выполнено $Eg(\xi) = \int_{\mathbb{R}} g(x)p(x)\mu(dx)$*

Определение 4.10. Пусть X – некоторое наблюдение с неизвестным распределением $P \in \{P_\theta : \theta \in \Theta\}$. Если $\forall \theta \in \Theta : P_\theta$ имеет плотность p_θ по одной и той же мере (либо мере Лебега, либо по считающей мере), то в этом случае $\{P_\theta : \theta \in \Theta\}$ называется доминируемым семейством.

Замечание. Для меры всегда будем использовать единое обозначение μ

4.6 Неравенство Рао-Крамера и эффективные оценки

Пусть X – наблюдение с неизвестным распределением $P \in \{P_\theta : \theta \in \Theta\}$ – доминируемое семейство с плотностью p_θ. Предположим, что выполнены следующие условия регулярности:

1. $A = \{x : p_\theta(x) > 0\}$ не зависит от параметра θ

2. Θ – открытый интервал на $\mathbb{R}$ (может быть бесконечный)

3. $\forall S(x) : E_\theta(S(X))^2 < \infty$ выполнено $\frac{\partial}{\partial\theta} E_\theta S(X) = E_\theta(S(X)\frac{\partial}{\partial\theta}\ln p_\theta(X))^2$

4. Интеграл $I_X(\theta) = E_\theta(\frac{\partial}{\partial\theta}\ln p_\theta(x))^2$ положителен и конечен $\forall\theta \in \Theta$

Определение 4.11. Случайная величина $U_\theta(X) = \frac{\partial}{\partial\theta}\ln p_\theta(X)$ называется вкладом в наблюдение X. $I_X(\theta) = E_\theta(U_\theta(X))^2$ называется количеством информации (по Фишеру), содержащейся в наюлюдении X

Теорема 4.1 (Неравенство Рао-Крамера)**.** *В условиях регулярности, если $\hat{\theta}(X)$ – несмещенная оценка $\tau(\theta)$, причем $E_\theta(\hat{\theta}(X))^2$ конечен $\forall\theta$. Тогда выполнено следующее неравенство: $D_\theta\hat{\theta}(X) \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}$*

Доказательство. Положим $S(x) \equiv 1$. В условии 3 получим $0 = E_\theta U_\theta(X)$. Возьмем теперь $S(x) = \hat{\theta}(X)$ в условии 3: $\frac{\partial}{\partial\theta}\hat{\theta}(X) = \tau(\theta) = E_\theta\hat{\theta}(X)U_\theta(X)$. То есть имеем: $\tau'(\theta) = E_\theta\hat{\theta}(X)U_\theta(X)$. Умножим первое равенство на $\tau(\theta)$ и вычтем из второго. Получим:

$\tau'(\theta) - 0 = E_\theta(\hat{\theta} - \tau(\theta))U_\theta(X)$ Применяем КБШ. Получаем, что $\tau'(\theta)^2 \leq D_\theta\hat{\theta} \cdot I_X(\theta)$ □

Следствие. *Если $\tau(\theta) = \theta$, то $D_{\hat{\theta}} \geq \frac{1}{I_X(\theta)}$*

Если $X = (X_1, \cdots, X_n)$ – выборка, то $I_X(\theta) = n \cdot i(\theta)$, где $i(\theta)$ – информация одного наблюдения. В этом случае $D_\theta\hat{\theta} = \Omega(\frac{1}{n})$

Определение 4.12. Оценки, в которых в неравенстве Рао-Крамера достигается равенство, называются эффективными.

4.7 Метод нахождения эффективных оценок

Лемма 4.2. *В неравенстве Рао-Крамера равенство достигается тогда и только тогда, когда $\hat{\theta}(X) - \tau(\theta) = c(\theta)U_\theta(X)$, где $c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$*

Доказательство. Равенство в неравенстве Коши-Буняковского достигается тогда и только тогда, когда случайные величины линейно зависимы почти наверное. Значит, $\hat{\theta}(X) - \tau(\theta) = c(\theta)U_\theta(X) + a(\theta)$. Чему равны эти $a(\theta)$ и $c(\theta)$? Берем математическое ожидание от обоих частей. Получаем, что $a(\theta) = 0$.

Теперь умножим обе части на вклад и возьмем математическое ожидание: $\tau'(\theta) = E_\theta(\hat{\theta}(X) - \tau(\theta))U_\theta(X) = c(\theta)E_\theta(U_\theta(X))^2 = I_X(\theta)$ □

Пример. Пусть $(X_1, \cdots, X_n) \sim Bin(1, \theta)$. $\theta \in (0, 1)$. Для каких $\tau(\theta)$ существуют эффективные оценки? Вычислить информацию Фишера одного наблюдения.

Рассмотрим плотность $p_\theta(x_1, \cdots, x_n) = \prod_{i=1}^n \theta^{x_i} \cdot (1-\theta)^{1-x_i} = \theta^{n\overline{X}} \cdot (1-\theta)^{n-n\overline{X}}$. Что такое вклад? $U_\theta(X) = \frac{\partial}{\partial\theta} \ln p_\theta = \frac{n\overline{X}}{\theta} - \frac{n-n\overline{X}}{1-\theta}$. Преобразуем последнее выражение: $\overline{X} - \theta = \frac{\theta(1-\theta)}{n} U_\theta(X)$. Получается, что мы нашли эффективную оценку $\overline{X}$ – это оценка параметра θ. Информация Фишера одного наблюдения $i(\theta) = \tau'(\theta) I_X(\theta)/n = \frac{1}{\theta(1-\theta)}$.

Кроме $\tau(\theta)$ можно эффективно оценить линейную функцию от $\tau(\theta)$

Замечание. Понятно, что эффективная оценка очень хороша – это видно из неравенства Рао-Крамера. Она является наилучшей оценкой $\tau(\theta)$ в среднеквадратическом подхоже в классе несмещённых оценок $\tau(\theta)$

4.8 Метод максимального правдоподобия

Определение 4.13. Пусть $(X_1, \cdots, X_n)$ – наблюдение из неизвестного распределения $P \in \{P_\theta : \theta \in \Theta\}$, где $\{P_\theta\}$ – доминируемое семейство с плотностью $p_\theta(X)$. Тогда функцией правдоподобия называется плотность, с подставленными туда результатми наблюдения: $f_\theta(X) = f_\theta(X_1, \cdots, X_n) = p_\theta(X)$. Если $X = (X_1, \cdots, X_n)$ – выборка из распределения с плотностью p_θ (то есть теперь p_θ – это неизвестная плотность X_i), то: $f_\theta(X) = f_\theta(X_1, \cdots, X_n) = \prod_{i=1}^n p_\theta(X_i)$

Определение 4.14. Оценка максимального правдоподобия (MLE, Maximum Likelihood Estimation) – это $\hat{\theta} = argmax_{\theta\in\Theta} f_\theta(X)$. Смысл этой оценки в том, мы живём в наиболее вероятном мире. То есть когда случайное событие происходит, то всегда выбирается то, у которого наибольшая вероятность.

Пример. $X = (X_1, \cdots, X_n)$ – выборка из равномерного распределения $U(0, \theta)$.

$f_\theta(X_1, \cdots, X_n) = \prod_{i=1}^n \frac{1}{\theta} \cdot I\{0 \le X_i \le \theta\} = \frac{1}{\theta^n} I\{0 \le X_{(1)}, X_{(n)} \le \theta\}$

Посмотрим на эту функцию, на отрезке $(0, +\infty)$ она возрастает в сторону нуля. Значит, логично взять минимально возможно значение θ, не нарушающее индикаторы. Это значение – $\hat{\theta} = X_{(n)}$

Пример. Пусть $X = (X_1, \cdots, X_n)$ – выборка из $\mathcal{N}(a, \sigma^2)$, причем оба параметра неизвестны.

$f_\theta(X) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_i-a)^2}{2\sigma^2}}$ Это гладкая функция, значит оценка по методу максимального правдоподобия, это решение системы:

$$\begin{cases} \frac{\partial}{\partial a} f_\theta(X_1, \cdots, X_n) = 0, \\ \frac{\partial}{\partial \sigma^2} f_\theta(X_1, \cdots, X_n) = 0 \end{cases}$$

Очень часто бывает выгодно сначала взять логарифм. $L_\theta(X) = \ln f_\theta(X)$ – логарифмическая функция правдоподобия.

Определение 4.15. Условия регулярности для MLE:

1. $X = (X_1, \cdots, X_n)$ – выборка растущего размера из неизветсного распределения $P \in \{P_\theta : \theta \in \Theta\}$

2. $\{P_\theta\}$ – доминируемое семейство с плотностью $p_\theta(X)$

3. $\forall \theta_1 \neq \theta_2 : P_{\theta_1} \neq P_{\theta_2}$

4. $A = \{x : p_\theta(x) > 0\}$ не зависит от θ.

Теорема 4.3 (экстремальное свойство правдоподобия)**.** *В условиях регулярности для MLE, если $f_\theta(X_1, \cdots, X_n)$ – функция правдоподобия. Тогда $\forall \theta \neq \theta_0 : P_{\theta_0}(f_{\theta_0}(X_1, \cdots, X_n) > f_\theta(X_1, \cdots, X_n)) \to 1$*

Доказательство. $f_{\theta_0}(X) > f_\theta(X) \Leftrightarrow \frac{f_{\theta_0}(X)}{f_{\theta(X)}} < 1 \Leftrightarrow \ln \frac{f_\theta}{f_{\theta_0}} < 0$ Умножим все на $\frac{1}{n}$.

$\frac{1}{n} \ln \frac{f_\theta}{f_{\theta_0}} = \frac{1}{n} \ln \prod_{i=1}^n \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} = \frac{1}{n} \sum_{i=1}^n \ln \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} \overset{\text{п.н.}}{\to} E_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)}$

Покажем, что $E_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1} < 0$.

$E_{\theta_0} \ln \frac{p_\theta(X_1}{p_{\theta_0}(X_1)} = \int_A \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x)\mu(dx) = \int_A \ln(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1) p_{\theta_0}(x)\mu(dx) \leq \int_A (\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1) p_{\theta_0}(x)\mu(dx) = 1 - 1 = 0$ Когда достигается равенство? Когда $P_{\theta_0}(\{x : p_\theta(x) = p_{\theta_0}(x)\}) = 1$. Но это не так. Поэтому неравенство строгое. Что и требовалось доказать. □

Следствие. *Если Θ конечно, то ОМП состоятельна.*

Доказательство. Из теоремы следует, что с вероятностью стремящейся к единице, ОМП будет равна истинному значению параметра. Тогда она состоятельна. □

Определение 4.16. Дополнительные условия регулярности:

1. Θ – открытый одномерный интервал

2. Пусть $p_\theta(x)$ дифференцируема по θ для всех $x \in A$

Теорема 4.4 (состоятельность решения уравнения правдоподобия)**.** *В расширенных условиях регулярности уравнение, называемое уравнением правдоподобия $\frac{\partial}{\partial \theta} \ln f_\theta(X_1, \cdots, X_n) = 0$ имеет решение, которое сходится по веротяности к истинному значению параметра.*

Доказательство. Пусть θ_0 – истинное значение параметра. Выберем такое $a > 0$, такое что $[\theta_0 - a, \theta_0 + a] \subset \Theta$. Тогда, по предыдущей теореме $P_{\theta_0}(f_{\theta_0 - a}(X_1, \cdots, X_n) < f_{\theta_0}(X_1, \cdots, X_n) > f_{\theta_0 + a}(X_1, \cdots, X_n)) \to 1$. Значит, с большой вероятностью внутри интервала $[\theta_0 - a, \theta_0 + a]$ найдется решение уравнения правдоподобия. Возьмем решение, ближайшее к истинному значению. $\hat{\theta}$ – оно самое. Тогда $\forall \varepsilon > 0$ в силу построения $P_{\theta_0}(|\hat{\theta} - \theta_0| < \varepsilon) \to 1$. Доказано. □

Вопросы:

1. $\hat{\theta}$ зависит от θ_0 (мы же выбрали ближайшее), которое мы хотим найти.

2. Как из множества корней отобрать ближайшее (при неизвестном θ_0).

3. Почему $\hat{\theta}$ – именно максимум, а не минимум.

Следствие. *Если в условиях теоремы $\forall X_1, \cdots, X_n$ существует ровно одно решение уравнения правдоподобия, то оно и является состоятельной оценкой параметра θ и с вероятностью стремящейся к единице это и есть оценка максимального правдоподобия.*

Доказательство. Состоятельность очевидна (следует из теоремы). Согласно доказательству теоремы $\forall \varepsilon > 0$ внутри отрезка $[\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ есть локальный максимум функции f_θ. Предположим, что есть другой глобальный максимум (или он не достигается), например вне этого отрезка. Но тогда между ними должен быть локальный минимум. Но локальный минмум тоже удовлетворяет уравнению правдоподобия, т.к функция дифференицируема. Но тогда решение у уравнения правдоподобия не единственно. Противоречие. $\square$

Определение 4.17. Ещё одни дополнительные условия регулярности:

1. Считаем, что плотность трижди непрерывно дифференцируема по θ $\forall x \in A$
2. $i(\theta) = E_\theta(\frac{\partial}{\partial\theta} \ln p_\theta(x))^2$ положительно и конечно для любого θ
3. $\forall \theta_0 \in \Theta : \exists c > 0 : \exists M(x) : \forall \theta \in (\theta_0 - c, \theta_0 + c) : |\frac{\partial^3}{\partial\theta^3} \ln p_\theta(x)| \leq M(x)$, причем $E_\theta M(X_i) < \infty$
4. Интеграл $\int_A p_\theta(x)\mu(dx)$ дважды дифферениируем по θ под знаком интеграла

Теорема 4.5 (асимптотическая нормальность состоятельного решения уравнения правдоподобия)**.** *В условиях регулярности $\forall$ состоятельное решение $\hat{\theta}$ является асимптотически нормальным и*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \frac{1}{i(\theta)})$$

Доказательство. Пусть θ_0 – истинное значение $\hat{\theta}_n \xrightarrow{P} \theta_0$. Разложим $\frac{\partial}{\partial\theta} L_\theta(x) = \frac{\partial}{\partial\theta} \ln f_\theta(x)$ в ряд тейлора в точке θ_0:

$$L'_\theta(X) = L'_{\theta_0}(X) + (\hat{\theta} - \theta_0)L''_{\theta_0}(X) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 L'''_\theta(X)(\overline{\theta})$$

, где $\overline{\theta}$ – какая-то промежуточная точка. По условию $L'_{\hat{\theta}_n}(x) = 0$. Перепишем это в таком виде:

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-L'_{\theta_0}(X)\frac{1}{\sqrt{n}}}{\frac{1}{n}(L''_{\theta_0}(X) + \frac{1}{2}(\hat{\theta} - \theta_0)L'''_{\overline{\theta}}(X)}$$

Числитель по ЦПТ сходится к 0 с асимптотической дисперсией $i(\theta)$

Также, по УЗБЧ $\frac{1}{n}L''_{\theta_0}(X) = \frac{1}{n}\sum_{k=1}^n \frac{\partial^2}{\partial\theta^2}(-\ln p_\theta(X_k))$ стремится по вероятности к $E\frac{\partial^2}{\partial\theta^2} - \ln p_\theta(X_k) = i(\theta)$ (проверятется ручками).

Осталось оценить $|\frac{1}{2n}(\hat\theta_n - \theta_0)L'''_{\overline{\theta}}(X)| = |(\hat\theta - \theta_0)\frac{1}{2}\frac{1}{n}\sum_{k=1}^n \frac{\partial^3}{\partial\theta^3} - \ln p_\theta(X_k)| \le |\frac{1}{n}\sum_{k=1}^n M(X_k)\frac{1}{2}(\hat\theta_n - \theta_0|$ сходится к нулю.

По лемме слуцкого, $\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{d} \frac{N(0,i(\theta))}{i(\theta)} = N(0, \frac{1}{i(\theta)})$ □

Следствие. *В условиях предыдущей теоремы, если $\forall x$: решение ОМП единственно, то ОМП асимптотически нормальная с наилучшей асимптотической дисперсией.*

Теорема 4.6 (Бахадур). *В условиях предыдущей теоремы, $\hat\theta_n$ – асимптотически нормальная оценка, $\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{d} N(0, \sigma(\theta)^2)$*

Следствие. *В условиях предыдущей теоремы, и предыдущего следствия, ОМП является наилучшей оценкой в асимптотическом подходе в классе оценок с непрерывной асимптотической диспресией.*

Определение 4.18. Оценка $\hat\theta_n$ называется асимптотически эффективной, если $\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{d} N(0, \frac{1}{i(\theta)}$

Теорема 4.7 (Эффективность ОМП). *Пусть выполнены условия регулярности, $\hat\theta$ – эффективная оценка θ. Тогда $\hat\theta$ – ОМП.*

Доказательство. Вспомним, что оценка эффективная, когда $\hat\theta(X) - \theta = \frac{1}{I(\theta)}\frac{\partial}{\partial\theta}f_\theta(X)$ Если $\theta > \hat\theta$, то f_θ убывает, если $\theta < \hat\theta$, то f_θ возрастает. Значит, $\theta = \hat\theta$ – это ОМП. □

4.9 Условное матожидание

Определение 4.19. Пусть $(\Omega, \mathcal{F}, P)$ – вероятностной пространство. Пусть ξ – случайная величина, а $\mathcal{G}$ – это сигма алгебра. ζ называется условным матожиданием ξ по алгебре $\mathcal{G}$ (обозначается $E(\xi|\mathcal{G})$), если:

1. ζ – является $\mathcal{G}$-измеримой случайной величиной ($\forall B \in \mathcal{B}(\mathbb{R}) : \zeta^{-1}(B) \in \mathcal{G}$)
2. $\forall G \in \mathcal{G} : E\xi I_G = E\zeta I_G$

Определение 4.20. ν – называется зарядом (или мерой со знаком), если:

1. $\nu : \mathcal{F} \mapsto \mathbb{R}$
2. $\forall A_1, A_2, \dots \in \mathcal{F}$ – непересекающихся, $\nu(\cup A_i) = \sum \nu(A_i)$

Теорема 4.8 (Радон, Никодим). *Пусть ν – заряд на $\mathcal{G}$. Тогда $\exists\zeta$ – случайная величина, такая что $\forall G \in \mathcal{G}$: $\int_G \zeta P(d\omega) = \nu(A)$, причем ζ является $\mathcal{G}$-измеримой.*

Теорема 4.9 (существование условного матожидания). *Пусть* $\xi : E\xi < \infty$. *Тогда* $\exists \zeta = E(\xi|\mathcal{G})$

Доказательство. Возьмем в качестве $\nu(A) = E\xi I_A$. Нетрудно убедится, что из теоремы Лебега о мажорируемой сходимости следует, что это действительно заряд. $|\xi \sum_{i=1}^{k} I_{A_i}| < \xi$. Но $\xi \sum_{i=1}^{k} I_{A_i} \overset{\text{п.н.}}{\to} \xi \sum_{i=1}^{\infty}$. По теореме Радона Никодима $\exists \zeta : E\zeta I_A = \nu(A)$ □

Пример. Берём $\mathcal{G}$ – разбиение на $G_1, G_2, \cdots$. Тогда $E(\xi|\mathcal{G}) = \sum_{i=1}^{\infty} c_i I_{G_i}$ из первого свойства, а из второго свойства $c_i = \frac{E\xi I_{G_i}}{P(G_i)}$

Свойства:

1. Условное матожидание единственно почти наверное. Действительно, пусть ζ_1, ζ_2 – два матожидания. $0 \leq E(\zeta_1 - \zeta_2)I_A \leq 0$. Значит, $\zeta_1 \leq \zeta_2$ почти наверное, $\zeta_1 \geq \zeta_2$ почти наверное.

2. Если ξ – $\mathcal{G}$-измерима, то $E(\xi|\mathcal{G}) = \xi$

3. $E(E(\xi|\mathcal{G})) = E\xi$

4. Если ξ не зависит от $\mathcal{G}$, то $E(\xi|\mathcal{G}) = E\xi$. Действительно, рассмотрим $\zeta = E\xi$. Оно удовлетворяет обоим свойствам условного матожидания.

5. $E(\alpha\xi_1 + \beta\xi_2|\mathcal{G}) = \alpha E(\xi_1|\mathcal{G}) + \beta E(\xi_2|\mathcal{G})$

5 Условное матожидание

Определение 5.1. Пусть $(\Omega, \mathcal{F}, P)$ – вероятностной пространство. Пусть ξ – случайная величина, а $\mathcal{G}$ – это сигма алгебра. ζ называется условным матожиданием ξ по алгебре $\mathcal{G}$ (обозначается $E(\xi|\mathcal{G})$), если:

1. ζ – является $\mathcal{G}$-измеримой случайной величиной ($\forall B \in \mathcal{B}(\mathbb{R}) : \zeta^{-1}(B) \in \mathcal{G}$)

2. $\forall G \in \mathcal{G} : E\xi I_G = E\zeta I_G$

Определение 5.2. ν – называется зарядом (или мерой со знаком), если:

1. $\nu : \mathcal{F} \mapsto \mathbb{R}$

2. $\forall A_1, A_2, \dots \in \mathcal{F}$ – непересекающихся, $\nu(\cup A_i) = \sum \nu(A_i)$

Теорема 5.1 (Радон, Никодим). *Пусть* ν *– заряд на* $\mathcal{G}$*. Тогда* $\exists \zeta$ *– случайная величина, такая что* $\forall G \in \mathcal{G}$*:* $\int_G \zeta P(d\omega) = \nu(A)$*, причем* ζ *является* $\mathcal{G}$*-измеримой.*

Теорема 5.2 (существование условного матожидания). *Пусть* $\xi : E\xi < \infty$. *Тогда* $\exists \zeta = E(\xi|\mathcal{G})$

Доказательство. Возьмем в качестве $\nu(A) = E\xi I_A$. Нетрудно убедится, что из теоремы Лебега о мажорируемой сходимости следует, что это действительно заряд. $|\xi \sum_{i=1}^{k} I_{A_i}| < \xi$. Но $\xi \sum_{i=1}^{k} I_{A_i} \overset{\text{п.н.}}{\to} \xi \sum_{i=1}^{\infty}$. По теореме Радона Никодима $\exists \zeta : E\zeta I_A = \nu(A)$ □

Пример. Берём $\mathcal{G}$ – разбиение на $G_1, G_2, \cdots$. Тогда $E(\xi|\mathcal{G}) = \sum_{i=1}^{\infty} c_i I_{G_i}$ из первого свойства, а из второго свойства $c_i = \frac{E\xi I_{G_i}}{P(G_i)}$

Свойства:

1. Условное матожидание единственно почти наверное. Действительно, пусть ζ_1, ζ_2 – два матожидания. $0 \leq E(\zeta_1 - \zeta_2)I_A \leq 0$. Значит, $\zeta_1 \leq \zeta_2$ почти наверное, $\zeta_1 \geq \zeta_2$ почти наверное.

2. Если ξ – $\mathcal{G}$-измерима, то $E(\xi|\mathcal{G}) = \xi$

3. $E(E(\xi|\mathcal{G})) = E\xi$

4. Если ξ не зависит от $\mathcal{G}$, то $E(\xi|\mathcal{G}) = E\xi$. Действительно, рассмотрим $\zeta = E\xi$. Оно удовлетворяет обоим свойствам условного матожидания.

5. $E(\alpha\xi_1 + \beta\xi_2|\mathcal{G}) = \alpha E(\xi_1|\mathcal{G}) + \beta E(\xi_2|\mathcal{G})$. Доказывается аналогично предыдущему пункту.

6. Если $\xi \leq \eta$, то $E(\xi|G) \leq E(\eta|G)$.

   Доказательство. Раз $\xi \leq \eta$, то $E(\xi I_A) \leq E(\eta I_A)$. Тогда из определения УМО $E(E(\xi|G)I_A) \leq E(E(\eta|G)I_A)$. Рассмотрим $\delta = E(\eta|G) - E(\xi|G)$. δ является G-измеримой. Кроме того $E(\delta I_A) \geq 0$. Тогда $\delta \geq 0$ почти наверное. □

7. $|E(\xi|G)| \leq E(|\xi||G|)$

8. Пусть $G_1 \subset G_2$. Тогда $E(E(\xi|G_1)|G_2) = E(\xi|G_1), E(E(\xi|G_2)|G_1) = E(\xi|G_1)$

   Доказательство. $E(\xi|G_1)$ является G_2 измеримой. Тогда $E(E(\xi|G_1)|G_2) = E(\xi|G_1)$. Для доказательства второго факта проверяем свойство 2 УМО. □

9. Пусть $\xi_n \overset{\text{п.н.}}{\to} \xi$, $\forall n : |\xi_n| < \eta, E\eta < \infty$. Тогда $E(\xi_n|G) \overset{\text{п.н.}}{\to} E(\xi|G)$

10. Пусть $E\xi, E\xi\eta$ конечны, η является G-измеримой случайной величиной. Тогда $E(\xi\eta|G) = \eta E(\xi|G)$

    Доказательство. Рассмотрим сначала $\eta = I_B, B \in G$. Тогда $E(\xi\eta I_A) = E(\xi I_{A\cap B}) = E(E(\xi|G)I_{A\cap B}) = E(\xi|G)\eta I_A)$. В силу линейности аналогичное верно и для $\eta = \sum_{i=1}^{k} c_i I_{A_i}$. Произвольную η аппроксимируем простыми G-измеримыми случайными величнами и воспользуемся теоремой Лебега (предыдущим свйоством) □

Теорема 5.3 (о наилучшем квадратичном прогнозе)**.** *Пусть G – сигма алгебра в $\mathcal{F}$. π_G – множество всех G-измеримых случайных величин с конечным вторым моментом. Если $E\xi^2 < \infty$, то $\min_{\eta\in\pi_G} E(\xi-\eta)^2 = E(\xi - E(\xi|G))^2$*

Доказательство. Пусть $\eta \in \pi_G$. Тогда $E(\xi-\eta)^2 = E(\xi - E(\xi|G) - (\eta - E(\xi|G)))^2 = E(\xi - E(\xi|G))^2 + E(\eta - E(\xi|G))^2 - 2E(\xi - E(\xi|G))(\eta - E(\xi|G))$. Покажем, что последнее слагаемое равно нулю. Действительно, $E(\xi - E(\xi|G))(\eta - E(\xi|G)) = E(E((\xi - E(\xi|G))(\eta - E(\xi|G))|G)) = E((\eta - E(\xi|G))E(\xi - E(\xi|G)|G)) = E((\eta - E(\xi|G))(E(\xi|G) - E(\xi|G))) = E0 = 0$. Тогда $E(\xi-\eta)^2 \geq E(\xi - E(\xi|G))^2$ □

6 Условное распределение

Определение 6.1. Пусть $A \in \mathcal{F}$, $\mathcal{G} \subset \mathcal{F}$ – под-сигма-алгебра. Определим $P(A|G) := E(I_A|G)$

Определение 6.2. Пусть η – случайная величина, $\mathcal{F}_\eta$ – сигма-алгебра, порожденная случайной величиной η. Тогда $E(\xi|\eta) := E(\xi|\mathcal{F}_\eta)$

Определение 6.3. Пусть ξ, η – случайные величины. Тогда $E(\xi|\eta = y)$ – это такая борелевская $\varphi(y)$, что $\forall B \in \mathcal{B}(\mathbb{R}) : E(\xi I\{\eta \in B\}) = \int_B \varphi(y) P_\eta(dy)$, где P_η – это распределение случайной величины η

Лемма 6.1. *Пусть $E|\xi| < \infty$, тогда $E(\xi|\eta = y)$ существует и единственно почти наверное.*

Доказательство. Следует из теоремы Радона-Никодима. □

Утверждение 6.2. $E(\xi|\eta = y) = \varphi(y) \Leftrightarrow E(\xi|\eta) = \varphi(\eta)$

Доказательство. Докажем сначала слева направо. Ясно, что $\varphi(\eta)$ является η-измеримой. Проверим интегральное свойство. Мы значем, что $E(\xi I\{\eta \in B\}) = \int_B \varphi(y) P_\eta(dy)$. Пусть $A \in \mathcal{F}_\eta$. Тогда $A = \{\eta \in B\}$ для некоторого $B \in \mathcal{B}(\mathbb{R})$. Тогда $E\xi I_A = E\xi I\{\eta \in B\} = \int_B \varphi(y) P_\eta(dy) = \int_{\mathbb{R}} \varphi(y) I_B P_\eta(dy) = E\varphi(\eta) I\{\eta \in B\}$. Последнее равенство выполнено в силу теоремы о замене переменной под интегралом Лебега. В обратную сторону утверждение доказывается аналогично (только теперь нам извество, что $\varphi(\eta)$ – это условное матожидание). □

Следствие. *ξ является η измеримой $\Leftrightarrow \xi = \varphi(\eta)$ для некоторой борелевской φ.*

Определение 6.4. Условным распределением случайной величины ξ при условии $\eta = y$ называется $P(\xi \in B|\eta = y)$ рассматриваемая как функция от B.

Утверждение 6.3. *При фиксированном y величина $P(\xi \in B|\eta = y)$ является вероятностной мерой P_η почти наверное на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*

Определение 6.5. Если условное распределение предствимо в виде $P(\xi \in B|\eta = y) = \int_B f_{\xi|\eta}(x|y)dx$, где функция $f_{\xi|\eta}(x|y)$ – это обычная неотрицательная функция двух переменных, то говорят, что $f_{\xi|\eta}$ является плотностью условного распределения.

Теорема 6.4. *Пусть $g(x)$ – борелевская функция, такая что $E|g(\xi)| < \infty$. Тогда при условии существования условной плотности $f_{\xi|\eta}$ можно записать следующее:*

$$E(g(\xi)|\eta = y) = \int_{\mathbb{R}} g(x) f_{\xi|\eta}(x|y)dx$$

Доказательство. Coming soon □

Доказательство. Нетрудно проверить, что $\int_A \varphi(x, y)dx$ является условным распределением $P(\xi \in A|\eta = y)$. Тогда $\forall B \in \mathcal{B}(\mathbb{R}) : E(I\{f \in A\}I\{\eta \in B\}) = P(\xi \in A, \eta \in B) = \int_{A\times B} f_{\xi,\eta}(x, y)dxdy = \int_B \left(\int_A f_{\xi,\eta}(x, y)f_\eta^{-1}(y)dx\right) f_\eta(y)dy = \int_B \left(\int_A \varphi(x, y)dx\right) P_\eta(dy) = \int_B P(\xi \in A|\eta = y)P_\eta(dy)$. Значит, $\varphi(x, y)$ – действительно условная плотность. □

Как искать условное математическое ожидание:

1. Находим совметсную плотность $f_{(\xi,\eta)}(x, y)$.
2. Находим плотность условия $f_\eta(y) = \int_{\mathbb{R}} f_{(\xi,\eta)}(x, y)dx$
3. Находим условную плотность $f_{\xi|\eta}(x|y) = \frac{f_{(\xi,\eta)}(x,y)}{f_\eta(y)}$
4. $\varphi(y) = E(g(\xi)|\eta = y) = \int_{\mathbb{R}} g(x)f_{\xi|\eta}(x|y)dx$
5. $E(g(\xi)|\eta) = \varphi(\eta)$

Утверждение 6.5. *В дискретном случае* $E(g(\xi)|\eta = y) = \sum_{x\in Im\xi} g(x)P(\xi = x|\eta = y)$

Доказательство. Упражнение. □

7 Баесовские оценки

В баесовском подходе мы хотим минимизировать $\int_\Theta R(\hat{\theta}(X), t)Q(dt)$, где Q – распределение вероятностей на Θ. Будем сичтать, что X имеет неизветсное распределение $P \in \{P_\theta : \theta \in \Theta\}$ – доминируемое семейтсво с плотностью $p_\theta(x)$. Будем считать, что Q имеет плотность $q(t)$ по мере ν.

Определение 7.1. Плотность $q(t)$ называется априорной плотностью q.

Определение 7.2. Величина $q(t|X) = \frac{q(t)p_t(X)}{\int_\Theta q(t)p_t(X)\nu(dt)}$ называется апостериорной плотностью Q.

Определение 7.3. Оценка $\hat{\theta} = E(\theta|X) = \int_\Theta t(t|X)\nu(dt)$ называется байсовской оценкой.

Теорема 7.1 (о байесовской оценке). *Байесовская оценка является наилучшей оценкой θ в байесовском подходе с квадратичной функцией потерь.*

Доказательство. Пусть $\Theta \subset \mathbb{R}$. Рассмотрим вероятностное пространство (Θ, B_Θ, Q), а $\theta(t) = t$. Тогда θ – это случайная величина на этом вероятностном пространстве с распределением Q.

Рассмотрим произведение вероятностных пространств $(\Theta\times\mathfrak{X}, \mathcal{B}_\Theta\times\mathcal{B}_\mathfrak{X}, \overline{P})$, где $\overline{P}$ имеет плотность $f(t,x) = q(t)p_t(x)$. Почему это плотность?

$\int_{\Theta\times\mathfrak{X}} f(t,x)\nu(dt)\mu(dx) = \int_\Theta q(t)\nu(dt)\int_\mathfrak{X} p_t(x)\mu(dx) = \int_\Theta q(t)\nu(dt) = 1$

Теперь введем $X : \Theta\times\mathfrak{X} \mapsto \mathfrak{X}$, где $X(t,x) = x$. Тогда X – случайная величина на этом вероятностном пространстве, а $U = (\theta X)^T$ – случайный вектор. $U(t,x) = (t,x)$, значит f – плотность U. Тогда $q(t)$ – плотность θ, $\int_\Theta q(t)p_t(x)\nu(dt)$ – плотность X, $p_t(x)$ – условная плотность X относительно θ, $q(t|x)$ – услованя плотность θ относительно X. Тогда Байесовская оценка – это условное матожидание $E(\theta|X)$.

Нам нужно минимизировать

$$\int_\Theta R(\hat{\theta}(X), t)Q(dt) = \int_\Theta E_t(\hat{\theta} - t)^2 q(t)\nu(dt) =$$
$$\int_\Theta\int_\mathfrak{X} (\hat{\theta}(X) - t)^2 p_t(x)\mu(dx)q(t)\nu(dt) =$$
$$\int_{\Theta\times\mathfrak{X}} (\hat{\theta}(X) - t)^2 f(t,x)\nu(dt)\mu(dx) = E(\hat{\theta}(X) - \theta)^2$$

То есть нам нужно минимизировать среднеквадратическое уклонение θ относительно функции $\hat{\theta}(X)$, то есть относительно X-измеримой функции. По теореме о наилучшем среднеквадратическом прогнозе минимум достигается при $\hat{\theta}(X) = E(\theta|X)$ □

8 Достаточные статистики и оптимальные оценки

Coming soon…

9 Доверительные интервалы

Определение 9.1. Доверительным интервалом уровня доверия γ называется такая пара статистик $(T_1(X), T_2(X)$, что $\forall\theta \in \Theta : P(T_1(X) \le \theta \le T_2(x)) \ge \gamma$. Если последняя вероятность $= \gamma$, то доверительный интервал называется точным. Также иногда бывает удобно рассматривать односторонние доверительные интервалы.

Определение 9.2. Область $S(X) \subset \Theta$ называется доверительной областью уровня доверия γ, если $P(\theta \in S(X)) \geq \gamma$.

9.1 Метод центральной статистики

Пусть $G(X, \theta)$ – одномерная случайная величина, распределение которой не зависит от θ. Такая $G(X, \theta)$ называется центральной статистикой.

Пусть распределение G извество, $\gamma_1, \gamma_2 \in (0,1)$, $\gamma_1 < \gamma_2$, $\gamma = \gamma_2 - \gamma_1$. Пусть g_1, g_2 – γ_1-, γ_2-квантили $G(X, \theta)$. Тогда $\forall \theta \in \Theta : P(g_1 < G(X, \theta) < g_2) \geq \gamma$. Значит $S(X) = \{\theta \in \Theta : g_1 < G(X, \theta) < g_2\}$ является доверительной областью уровня доверия γ

Пример. Рассмотрим $N(\theta, 1)$. Построим доверительный интервал для θ.

$X_i - \theta \ N(0,1)$. Тогда $\sum_{i=1}^n (X_i - \theta) \ N(0, n)$. Тогда $\frac{1}{\sqrt{n}} (\sum_{i=1}^n (X_i - \theta)) = \sqrt{n}(\overline{X} - \theta) \ N(0,1)$ – центральная статистика.

Пусть $\gamma_2 - \gamma_1 = \gamma$. u_p – квантиль уровня p распределения $N(0,1)$. Тогда $P(u_{\gamma_1} < \sqrt{n}(\overline{X} - \theta) < u_{\gamma_2}) = \gamma$. Тогда $(\overline{X} - u_{\gamma_1}/\sqrt{n}, \overline{X} - u_{\gamma_2}/\sqrt{n})$ – довирительный интервал для θ, причем его длина стремится к нулю при n стремяшемся к бесконечности.