

Математическая статистика

18 июня 2020 г.

Содержание

1	Постановка задач математической статистики	4
1.1	Задачи теории вероятностей	4
1.2	Задачи математической статистики	4
2	Частота как оценка вероятности события и её свойства. Построение доверительного интервала для вероятности события на базе асимптотической нормальности частоты.	6
3	Постановка выборочной статистической модели. Точечная оценка параметра и характеристики.	8
4	Функции потерь и функции риска, состоятельность оценки характеристики, достаточное условие для состоятельности оценки.	9
5	Вид квадратичного риска в случае одномерной характеристики.	11
6	Постановка задачи доверительного оценивания, доверительный интервал.	12
7	Определение несмещенности и асимптотической нормальности оценки характеристики. Построение доверительного интервала для характеристики на базе асимптотической нормальности ее оценки.	13
8	Постановка задачи проверки гипотез	14
9	Ошибки первого и второго рода и их вероятности как критерий качества критерия (теста) проверки гипотез. Подход Неймана-Пирсона.	15
10	Асимптотический вариант задачи проверки гипотез. Состоятельный тест асимптотического уровня значимости α .	16
11	Эмпирическая функция распределения (ЭФР). Построение, свойства ЭФР при фиксированном значении аргумента (использовать свойства частоты).	17

12 Свойства ЭФР в целом. Расстояние Колмогорова, Смирнова. Теоремы Гливленко-Кантелли, Колмогорова, Мизеса-Смирнова. Построение доверительной полосы для функции распределения.	18
13 Критерии согласия Колмогорова и Мизеса-Смирнова.	20
13.1 Критерий согласия Колмогорова	20
13.2 Критерий Мизеса-Смирнова	20
13.3 Прикладной алгоритм	21
14 Выборочный метод построения оценок одномерных характеристик. Асимптотическая нормальность оценки. Построение асимптотического доверительного интервала на базе асимптотической нормальности.	22
14.1 Описание выборочного метода	22
14.2 Асимптотическая нормальность, свойства асимптотической нормальности оценок	22
15 Основные выборочные оценки и их свойства. Выборочное математическое ожидание. Выборочная дисперсия. Выборочные моменты. Выборочные медиана и квантили. Выборочные оценки ковариации и коэффициента корреляции.	24
15.1 Выборочное среднее / М.О.	24
15.2 Выборочная дисперсия	25
15.3 Несмещенная выборочная дисперсия	25
15.4 Выборочные моменты	26
15.4.1 Выборочные начальные моменты	26
15.4.2 Выборочные центральные моменты	26
15.5 Выборочная медиана	26
15.6 Выборочная ковариация и корреляция	27
15.6.1 Выборочная ковариация	27
15.7 Выборочная корреляция	28
16 Гистограмма как оценка плотности распределения. Статистические свойства гистограммы. Теорема Пирсона. Критерий хи-квадрат для проверки гипотезы о виде распределения генеральной совокупности	29
16.1 Построение	29
16.2 Статистические свойства гистограммы	30
16.3 Критерий хи-квадрат	31
16.3.1 Дискретная случайная величина	31
16.3.2 Критерий хи-квадрат для случайной величины общего вида	32
17 Метод моментов и его свойства.	33
17.1 Идея метода подстановки	33
17.2 Метод моментов	33
18 Метод максимального правдоподобия и его свойства.	35

19 О сравнении качества оценок. Свойства функции правдоподобия (одномерный параметр). Неравенство Рао-Крамера и эффективные оценки.	37
19.1 О сравнении качества оценок	37
19.1.1 Минимаксный подход	37
19.1.2 Асимптотически минимаксные оценки	37
19.2 Свойства функции правдоподобия (одномерный параметр)	38
19.3 Неравенство Рао-Крамера и эффективные оценки	38
20 Примеры построения наиболее мощных и равномерно наиболее мощных тестов.	40
20.1 Пример 1	40
21 Доверительное оценивание и проверка гипотез на основе оценок максимального правдоподобия.	41
21.1 Доверительное оценивание	41
21.2 Проверка гипотез	41
22 Общая линейная модель или задачи регрессии	43
23 Простейшие случайные процессы. Общие определения. Примеры. Моменты.	45
24 Цепи Маркова. Марковская зависимость. Переходные вероятности. Предельные вероятности. Схемы блужданий.	47
24.1 Марковская зависимость	47
24.2 Предельные вероятности	48
24.3 Схемы блужданий	51
24.3.1 Блуждание по отрезку с поглощением на концах	51
24.3.2 Блуждание по отрезку с отражением на концах	52

1 Постановка задач математической статистики

Сравним задачи теории вероятностей и математической статистики

1.1 Задачи теории вероятностей

Заданы:

- Вероятностное пространство $\langle \Omega, \Sigma, P \rangle$.
- Случайная величина $X : \Omega \rightarrow \mathbb{R}^n$.

Требуется получить различного рода характеристики величины X и величин, получающихся из X .

1.2 Задачи математической статистики

Определение. *Статистическим экспериментом* называется четверка

$$\langle \mathcal{X}, \mathcal{A}, P_\theta, \Theta \rangle.$$

Здесь:

- \mathcal{X} – множество наблюдений.
- \mathcal{A} – σ -алгебра подмножеств \mathcal{X} .
- P_θ – известная с точностью до неизвестного параметра θ вероятностная мера – закон распределения наблюдаемых данных.
- Θ – множество допустимых значений неизвестного параметра, то есть $\theta \in \Theta$.

Задачей математической статистики является получение той или иной информации о законе распределения наблюдаемых данных $P = P_\theta$.

Определение. *Статистикой* называется измеримая функция

$$f : \mathcal{X} \rightarrow A.$$

Для произвольного A .

Определение. Пусть

$$\bar{X} = \langle X_1, \dots, X_n \rangle.$$

Где $X_i \sim X$ – одинаково распределенные случайные величины. Соответствующая модель называется *моделью независимой однородной выборки*.

Определение. *Гипотезой* H называется подмножество Θ :

$$H \subseteq \Theta.$$

Перечислим некоторые задачи математической статистики.

- Оценивание параметра θ или какой-либо функции $g(\theta)$, то есть построение статистики $\hat{g}: \mathcal{X} \rightarrow \Theta$. Оценивание может быть:
 - *точечным*, то есть указание численной оценки $g(\theta)$
 - *доверительным*, то есть указание множества, с фиксированной вероятностью содержащего $g(\theta)$
- Проверка гипотез. Пусть имеется разбиение Θ на гипотезы: $\Theta = \bigsqcup_{n \in N} H_n$. Тогда проверкой гипотезы назовем построение *теста (критерия)*, то есть отображения

$$\varphi: \mathcal{X} \rightarrow N.$$

Которое по наблюдению выдает номер гипотезы, которому это наблюдение “соответствует”.

Естественно, перечисленные задачи можно оценивать с точки зрения качества. В этом смысле всегда требуется с точки зрения какой-либо метрики построить “лучшую” оценку.

2 Частота как оценка вероятности события и её свойства. Построение доверительного интервала для вероятности события на базе асимптотической нормальности частоты.

Теорема 2.1. (Яков, Бернулли)

Пусть имеется $\xi_i \sim \xi$ – последовательность одинаково распределенных и попарно независимых случайных величин. Пусть

$$\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i = \frac{k_n}{n}.$$

Тогда

$$\bar{\xi}_n \xrightarrow[n \rightarrow +\infty]{} p.$$

Теорема 2.2. (Центральная предельная теорема, простейший вариант)

Пусть случайные величины $X_i \sim X$ независимы и одинаково распределены, причем $\exists E(X), D(X)$. Тогда для случайной величины

$$Y_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sigma(\bar{X}_n)}.$$

Верно:

$$F_{Y_n} \xrightarrow[\mathbb{R}]{} F_{N(0,1)}.$$

Теорема 2.3. (Свойства частоты как оценки p)

Пусть $\xi \sim B(p)$. Тогда

$$\hat{p} = \frac{k_n}{n}$$

Является несмещенной асимптотически нормальной оценкой p , то есть

$$E(\hat{p}) = p,$$

$$\sqrt{n} \cdot (\hat{p} - p) = Y_n \xrightarrow{P_{n,\theta}} Y \sim N(0, \Delta^2(p)), \Delta^2(p) = p(1-p).$$

Доказательство.

- Покажем несмещенность:

$$E(\hat{p}) = E\left(\frac{k_n}{n}\right) = \frac{1}{n}np = p.$$

- Асимптотическая нормальность с нормирующим множителем $\Delta^2(p) = p(1-p)$ следует непосредственно из центральной предельной теоремы.

■

На базе асимптотической нормальности можно построить доверительный интервал. Проделаем это на примере частоты. Выпишем определение асимптотической нормальности:

$$Y_n = \frac{\sqrt{n} \cdot (\hat{p} - p)}{\sqrt{p(1-p)}} \rightarrow N(0, 1).$$

Это буквально означает:

$$P_{n,\theta}(Y_n < t) \rightarrow F_{N(0,1)}(t).$$

Раскроем определение Y_n , возьмем его по модулю и воспользуемся квантилью:

$$P_{n,\theta} \left(\left| \frac{\sqrt{n} \cdot (\hat{p} - p)}{\sqrt{p(1-p)}} \right| < t_\gamma \right) \rightarrow \gamma \iff P_{n,\theta} \left(\frac{\sqrt{p(1-p)}}{\sqrt{n}} t_\gamma + \hat{p} > p > -\frac{\sqrt{p(1-p)}}{\sqrt{n}} t_\gamma + \hat{p} \right) \rightarrow \gamma.$$

Здесь $\gamma = P(|\xi| < t_\gamma)$, $\xi \sim N(0, 1)$.

3 Постановка выборочной статистической модели. Точечная оценка параметра и характеристики.

Определение. Напомним, что *точечной оценкой* параметра θ или какой-либо функции $g(\theta)$ называют численную оценку этой величины.

Пусть \hat{g} является некоторой точечной оценкой $g = g(\theta)$.

Определение. \hat{g} называется *несмещенной*, если $E(\hat{g}) = g(\theta)$.

Определение. При асимптотическом подходе оценка \hat{g} называется *состоятельной*, если $\hat{g} \xrightarrow[p]{} g(\theta)$ при $n \rightarrow \infty$.

Определение. \hat{g}_n называется *асимптотически нормальной*, если

$$\frac{\sqrt{n}(\hat{g}_n - g(\theta))}{\sigma(g(\theta))} \xrightarrow{P_{n,\theta}} N(0, 1).$$

Определение. \hat{g}_n называется *эффективной* в классе оценок K , если для любой другой оценки $\hat{g}_n^* \in K$ имеет место неравенство:

$$E(\hat{g}_n - g(\theta))^2 \leq E(\hat{g}_n^* - g(\theta))^2.$$

4 Функции потерь и функции риска, состоятельность оценки характеристики, достаточное условие для состоятельности оценки.

Определение. Оценкой $g(\theta)$ называется статистика вида

$$\hat{g}: \mathcal{X} \rightarrow g(\Theta).$$

Определение. Пусть $\hat{g}(\theta)$ – оценка $g(\theta)$. Тогда функцией потерь называется неотрицательная функция $l(\hat{g}, g(\theta))$, характеризующая “близость” оценки к настоящему значению.

Замечание. Обычно в качестве функции потерь рассматривают функцию вида

$$l(\hat{g}, g(\theta)) = \omega(\|\hat{g}, g(\theta)\|).$$

Здесь ω – неотрицательная монотонно возрастающая функция, $\omega(0) = 0$.

Замечание. l является случайной величиной.

Определение. Риском называется функция

$$R(\hat{g}, \theta) \stackrel{\text{def}}{=} E_{\theta}(l(\hat{g}, g(\theta))).$$

Замечание. Риск – функция параметра θ и способа оценивания \hat{g} .

Опишем самые важные для нас виды функции потерь и риска.

Определение. Определим функцию потерь индикатором отклонений:

$$l^{\delta}(\hat{g}, g(\theta)) = \omega^{\delta}(\|\hat{g}, g(\theta)\|).$$

Где

$$\omega(t) = \mathbb{1}_{\delta}(t) = \begin{cases} 0, & t < \delta \\ 1, & t \geq \delta \end{cases}.$$

Соответствующий риск будет вероятностью отклонения:

$$R^{\delta}(\hat{g}, \theta) = E_{\theta}(l^{\delta}(\hat{g}, g(\theta))) = 0 \cdot P_{\theta}(\|\hat{g}, g(\theta)\| < \delta) + 1 \cdot P_{\theta}(\|\hat{g}, g(\theta)\| \geq \delta) = P_{\theta}(\|\hat{g}, g(\theta)\| \geq \delta).$$

Определение. При асимптотическом подходе оценка называется *состоятельной*, если

$$\forall \delta > 0 \quad R^{\delta}(\hat{g}_n, \theta) = P_{n, \theta}(\|\hat{g}_n, g(\theta)\| \geq \delta) \xrightarrow{n \rightarrow +\infty} 0.$$

Или, что то же самое:

$$\hat{g}_n \xrightarrow[n \rightarrow +\infty]{P_{n, \theta}} g(\theta).$$

Определение. Квадратичной функцией потерь называется функция

$$l_2(\hat{g}, g(\theta)) = \|\hat{g}, g(\theta)\|^2.$$

Соответствующий ей риск называется квадратичным:

$$R_2(\hat{g}, \theta) = E_\theta(\|\hat{g}, g(\theta)\|^2).$$

Теорема 4.1. (Достаточное условие для состоятельности оценки)

В случае одномерной оценки $R_2(\hat{g}_n, \theta) \xrightarrow{n \rightarrow +\infty} 0 \implies$ оценка состоятельна.

Доказательство.

$$\begin{aligned} \forall \delta > 0 \quad R^\delta(\hat{g}_n, \theta) &= P(\|\hat{g}_n - g(\theta)\| \geq \delta) = P(\|\hat{g}_n - g(\theta)\|^2 \geq \delta^2) \\ &\leq \frac{E_\theta(\|\hat{g}_n - g(\theta)\|^2)}{\delta^2} = \frac{R_2(\hat{g}_n, \theta)}{\delta^2} \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

■

5 Вид квадратичного риска в случае одномерной характеристики.

Определение. Смещением оценки называется величина

$$b(\hat{g}, \theta) = g(\theta) - E_{\theta}(\hat{g}).$$

Определение. Оценка называется несмещенной, если $b(\hat{g}, \theta) = 0$.

Теорема 5.1. $R_2(\hat{g}, \theta) = D_{\theta}(\hat{g}) + b^2(\hat{g}, \theta)$.

Доказательство.

$$\begin{aligned} R_2(\hat{g}, \theta) &= E_{\theta}(\|\hat{g} - g(\theta)\|^2) = E_{\theta}(\hat{g} - E_{\theta}(\hat{g}) - (g(\theta) - E_{\theta}(\hat{g})))^2 \\ &= E_{\theta}(\hat{g} - E_{\theta}(\hat{g}))^2 + (g(\theta) - E_{\theta}(\hat{g}))^2 - \underbrace{2(g(\theta) - E_{\theta}(\hat{g}))(E_{\theta}\hat{g} - E_{\theta}\hat{g})}_0 \\ &= D_{\theta}(\hat{g}) + b^2(\hat{g}, \theta). \end{aligned}$$

■

Следствие 5.2. Для одномерных несмещенных оценок квадратичный риск в точности равен дисперсии оценки:

$$R_2(\hat{g}, \theta) = D_{\theta}(\hat{g}).$$

6 Постановка задачи доверительного оценивания, доверительный интервал.

При оценивании параметров или характеристик распределений мы в качестве результата получаем числовое значение $\hat{g}(X) \in g(\Theta)$. Такой способ оценивания мы называем *точечной оценкой*. Заранее не понятно, насколько результат соответствует действительности. Для того, чтобы можно было оценивать качество результата, нужно предъявлять не точку, а подмножество в $g(\Theta)$, содержащее в некотором смысле наиболее подходящие значения.

Задача доверительного оценивания ставится следующим образом: задана величина $\gamma \in (0, 1)$, называемая *уровнем надежности*. По заданному наблюдению X и значению надежности требуется построить доверительную область надежности.

Определение. Доверительной областью надежности называется $\tilde{G}_\gamma \subseteq G = g(\Theta)$, обладающая свойством:

$$\forall \theta \in \Theta P_\theta(g(\theta) \in \tilde{G}_\gamma) \geq \gamma.$$

То есть множество, с достаточной вероятностью содержащее оцениваемую величину.

Определение. В случае одномерной оценки чаще всего доверительные области надежности выбирают в виде промежутков, которые называются *доверительными интервалами*.

Определение. В асимптотическом случае (когда имеется последовательность оценок и статистических экспериментов) последовательность *асимптотических областей надежности* $\tilde{G}_{n,\gamma}$ задается условием:

$$\forall \theta \in \Theta \lim P_{n,\theta}(g(\theta) \in \tilde{G}_{n,\gamma}) \geq \gamma.$$

Определение. Аналогично задается последовательность асимптотических доверительных интервалов в случае одномерной характеристики.

7 Определение несмещенности и асимптотической нормальности оценки характеристики. Построение доверительного интервала для характеристики на базе асимптотической нормальности ее оценки.

Определение. Напомним, оценка называется *несмещенной*, если

$$b(\hat{g}, \theta) = g(\theta) - E_{\theta}(\hat{g}) = 0.$$

Определение. Последовательность оценок \hat{g}_n называется *асимптотически нормальной*, если

$$\sqrt{n} \cdot (\hat{g}_n - g(\theta)) = Y_n \xrightarrow{P_{n,\theta}} Y \sim N(0, \Delta^2(\theta)).$$

Определение. Величина $\Delta(\theta)$ из определения асимптотически нормальной оценки называется *нормирующим множителем*.

Замечание. Определение асимптотически нормальной оценки можно переписать так:

$$\frac{\sqrt{n} \cdot (\hat{g}_n - g(\theta))}{\Delta(\theta)} \xrightarrow{P_{n,\theta}} Y \sim N(0, 1).$$

На базе асимптотической нормальности можно построить доверительный интервал. Выпишем определение асимптотической нормальности:

$$Y_n = \frac{\sqrt{n} \cdot (\hat{g} - g(\theta))}{\Delta(\theta)} \rightarrow N(0, 1).$$

Это буквально означает:

$$P_{n,\theta}(Y_n < t) \rightarrow F_{N(0,1)}(t).$$

Раскроем определение Y_n , возьмем его по модулю и воспользуемся квантилью:

$$P_{n,\theta}\left(\left|\frac{\sqrt{n} \cdot (\hat{g} - g(\theta))}{\Delta(\theta)}\right| < t_{\gamma}\right) \rightarrow \gamma \iff P_{n,\theta}\left(\frac{\Delta(\theta)}{\sqrt{n}}t_{\gamma} + \hat{g} > g(\theta) > -\frac{\Delta(\theta)}{\sqrt{n}}t_{\gamma} + \hat{g}\right) \rightarrow \gamma.$$

Здесь $\gamma = P(|\xi| < t_{\gamma})$, $\xi \sim N(0, 1)$.

8 Постановка задачи проверки гипотез

Определение. *Гипотезой* называется множество предполагаемых зафиксированных значений некоторого подмножества неизвестных параметров:

$$H : \theta \in \Theta_H \subseteq \Theta.$$

Определение. Гипотезу называют *простой*, если $|H| = 1$.

Определение. Гипотезу называют *сложной*, если $|H| > 1$.

Определение. Гипотезами *согласия* называют набор из двух гипотез: основной H_0 и альтернативы H_1 , причем $H_0 = \overline{H_1}$.

Определение. Правило принятия или отклонения основной гипотезы H_0 называют *тестом (критерием)* проверки гипотезы:

$$\varphi(X) : X_n \rightarrow \{0, 1\}.$$

При этом:

- $X_{n,0}$ называют *допустимым множеством*.
- $X_{n,1}$ называют *критическим множеством*.
- $X_{n,0} \sqcup X_{n,1} = X_n$.

Определение. Случайная величина $L(\bar{X}) : X_n \rightarrow \mathbb{R}$ называется *тестовой статистикой*, если она служит порогом для правила принятия или отклонения основной гипотезы:

$$\varphi(\bar{X}) = \begin{cases} 0, & L(\bar{X}) < T(H_0) \\ 1, & L(\bar{X}) \geq T(H_0) \end{cases}.$$

Где T называют *порогом принятия решения*.

9 Ошибки первого и второго рода и их вероятности как критерий качества критерия (теста) проверки гипотез. Подход Неймана-Пирсона.

Определение. *Ошибкой I рода* называют отклонение основной гипотезы, в то время как она была верна.

Определение. *Ошибкой II рода* называют принятие основной гипотезы, в то время как она не была верна.

Определение. α называют *вероятностью ошибки I рода*:

$$\alpha(\varphi, \theta) \stackrel{\text{def}}{=} P_{\theta}(X_{n,1}), \theta \in \Theta_{H_0}.$$

Определение. *Уровнем значимости теста* называют верхнюю границу вероятности ошибки I рода по всем возможным наблюдаемым значениям неизвестных параметров, отвечающих основной гипотезе:

$$\alpha(\varphi) \stackrel{\text{def}}{=} \sup_{\theta \in \Theta_{H_0}} \alpha(\varphi, \theta).$$

Определение. β называют *вероятностью ошибки II рода*:

$$\beta(\varphi, \theta) \stackrel{\text{def}}{=} P_{\theta}(X_{n,0}), \theta \in \Theta_{H_1}.$$

Определение. *Мощностью теста* называют следующую величину:

$$\gamma(\varphi, \theta) \stackrel{\text{def}}{=} 1 - \beta(\varphi, \theta).$$

Подход Неймана-Пирсона.

Зафиксируем $\alpha \in (0, 1)$ (обычно выбирают малое значение). Будем считать это значение минимальной допустимой величиной ошибки I рода (*допустимый уровень значимости*).

Рассмотрим множество всех тестов таких, что:

$$\overline{\Phi}_{\alpha} = \{\varphi = \varphi(x) \mid \alpha(\varphi) \leq \alpha\}.$$

Среди этих тестов выбирается тест с минимальным значением β .

В асимптотических задачах ограничения накладываются на предельные значения.

10 Асимптотический вариант задачи проверки гипотез. Состоятельный тест асимптотического уровня значимости α .

При асимптотическом подходе последовательность тестов $\varphi = \varphi_n$ называют просто тестом и проводят исследование асимптотических (предельных) свойств тестов $\varphi = \{\varphi_n\}$ при $n \rightarrow \infty$.

Определение. Тест $\varphi = \{\varphi_n\}$ имеет асимптотический уровень значимости $\alpha(\varphi)$, $\alpha(\varphi) \in [0, 1]$, если:

$$\alpha_n(\varphi_n) = \sup_{\theta \in \Theta_{H_0}} \alpha(\varphi_n, \theta) \rightarrow \alpha(\varphi), \quad n \rightarrow \infty.$$

При использовании подхода Неймана-Пирсона в асимптотическом варианте ограничение накладывается на асимптотический уровень значимости: $\alpha(\varphi) = \alpha$.

Определение. При асимптотическом подходе тест $\varphi = \{\varphi_n\}$ называется *состоятельным*, если для любого $\theta \in \Theta_{H_1}$:

$$\beta(\varphi_n, \theta) \xrightarrow{n \rightarrow \infty} 0.$$

Определение. Мерой близости альтернативы $\theta \in \Theta_{H_1}$ и гипотезы H_0 называют следующую величину:

$$\rho(\theta, \Theta_{H_0}) = \inf_{\theta_0 \in \Theta_{H_0}} \|\theta - \theta_0\|.$$

Определение. Тест $\varphi = \{\varphi_n\}$ называется \sqrt{n} -состоятельным, если:

$$\beta(\varphi_n, \theta_n) \xrightarrow{n \rightarrow \infty} 0$$

Для такой последовательности $\theta_n \in \Theta_{H_1}$, что:

$$\sqrt{n}\rho(\theta_n, \Theta_{H_0}) \rightarrow \infty.$$

11 Эмпирическая функция распределения (ЭФР). Построение, свойства ЭФР при фиксированном значении аргумента (использовать свойства частоты).

Определение. Эмпирической функцией распределения (ЭФР) называют следующую оценку функции распределения генеральной совокупности:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, t]}.$$

Иными словами, значение ЭФР в точке t равно отношению числа наблюдений, меньших t , к их общему числу n .

Свойства ЭФР:

1. ЭФР кусочно-постоянна.
2. Скачки ЭФР имеют вид $\frac{k}{n}$ для некоторого $k \in (1; n)$.
3. Область принимаемых значений: $[0; 1]$.
4. Частота может служить как оценка функции распределения генеральной совокупности. При фиксированном $t = t_0$:

$$F_x(t_0) \approx F_n(t_0) = \xi_1 + \dots + \xi_n = \frac{k_n}{n} - \text{частота}.$$

5. $F_n(t)$ является состоятельной оценкой:

$$F_n(t_0) = \bar{\xi}_n : F_n(t_0) \xrightarrow{p=1} F_x.$$

6. $F_n(t)$ является асимптотически нормальной оценкой. Свойства частоты по типу нормальности рассмотрены в секции 2.

12 Свойства ЭФР в целом. Расстояние Колмогорова, Смирнова. Теоремы Гливленко-Кантелли, Колмогорова, Мизеса-Смирнова. Построение доверительной полосы для функции распределения.

Со свойствами ЭФР можно ознакомиться в предыдущем разделе.

Определение. Расстояние Колмогорова:

$$\rho_{\infty}(F_n, F_x) = \sup_t |F_n(t) - F_x(t)|.$$

Определение. Расстояние Смирнова:

$$\rho_2^2(F_n, F_x) = \int_{\mathbb{R}} (F_n(t) - F_x(t))^2 dF_x(t).$$

Теорема 12.1. (Гливленко-Кантелли)

Пусть \mathcal{F} – множество функций распределения. Тогда $\forall F_x(t) \in \mathcal{F}$ с вероятностью 1 справедливо предельное неравенство:

$$\rho_{\infty}(F_n, F_x) \xrightarrow{n \rightarrow \infty} 0.$$

Так как $\rho_2 \leq \rho_{\infty}$, то же верно для ρ_2 .

Замечание. $F_n(t)$ – состоятельная оценка $F_x(t)$ в расстояниях Колмогорова и Смирнова.

Пусть \mathcal{F}_c – множество всех непрерывных функций распределения.

Теорема 12.2. (Колмогоров)

$$P_{n,F}(\sqrt{n}\rho_{\infty}(F_n, F_x) < u) \xrightarrow{n \rightarrow \infty} \mathcal{K}(u) = \begin{cases} 0, & u = 0 \\ \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2(ju)^2}, & u > 0 \end{cases}.$$

Теорема 12.3. (Мизес, Смирнов)

$$P_{n,F}(\sqrt{n}\rho_2^2(F_n, F_x) < u) \xrightarrow{n \rightarrow \infty} \mathcal{S}(u),$$

где $\mathcal{S}(u)$ есть функция распределения следующей случайной величины:

$$\mathcal{U} = \sum_{j=1}^{\infty} \frac{\xi_j^2}{j^2 \pi^2}, \quad \xi_j \sim N(0, 1), \text{ независимые.}$$

Замечание. Используя теорему Колмогорова, можно построить доверительную полосу для функции распределения.

Определение. Доверительной полосой называют часть плоскости, в которую с надежностью γ попадает функция распределения генеральной совокупности:

$$\begin{cases} F_n^-(t) = \max\left(0, F_n(t) - \frac{u_{\gamma}}{\sqrt{n}}\right) \\ F_n^+(t) = \min\left(1, F_n(t) + \frac{u_{\gamma}}{\sqrt{n}}\right) \end{cases}, \quad \text{где } \mathcal{K}(u_{\gamma}) = \gamma.$$

Утверждение 12.4.

$$P_x(F_n^-(t) \leq F_x(t) \leq F_n^+(t)) \xrightarrow{n \rightarrow \infty} \gamma.$$

Доказательство. $0 \leq F_x(t) \leq 1$ всегда, тогда:

$$\begin{aligned} P_x(F_n^-(t) \leq F_x(t) \leq F_n^+(t)) &= P_x\left(F_n(t) - \frac{u_\gamma}{\sqrt{n}} \leq F_x(t) \leq F_n(t) + \frac{u_\gamma}{\sqrt{n}}\right) \stackrel{\forall t}{=} \\ &\stackrel{\forall t}{=} P_x(\sqrt{n}|F_x(t) - F_n(t)| \leq u_\gamma) \stackrel{\forall t}{=} \\ &\stackrel{\forall t}{=} P_x\left(\sqrt{n} \sup_t |F_x(t) - F_n(t)| \leq u_\gamma\right) \xrightarrow{\text{th. Колмогорова}} \mathcal{K}(u_\gamma) = \gamma. \end{aligned}$$

■

13 Критерии согласия Колмогорова и Мизеса-Смирнова.

Пусть $F_0(t)$ – заданная непрерывная функция распределения.

Поставим задачу проверки согласия:

$$H_0 \equiv (F_x(t) \equiv F_0(t)).$$

13.1 Критерий согласия Колмогорова

Определим тестовую статистику:

$$L(\bar{X}) = \sqrt{n}\rho_\infty(F_0, F_n).$$

По th. Колмогорова:

$$P(L(\bar{X}) < z) \xrightarrow{n \rightarrow \infty} \mathcal{K}(z),$$

где \mathcal{K} – распределение Колмогорова. Тогда порогом принятия решения при уровне значимости α является квантиль распределения Колмогорова порядка $1 - \alpha$ (далее $u_{1-\alpha}$).

Таким образом, определим тест:

$$\varphi(\bar{X}) = \begin{cases} 0, & \sqrt{n}\rho_\infty(F_0, F_n) < u_{1-\alpha} \\ 1, & \sqrt{n}\rho_\infty(F_0, F_n) \geq u_{1-\alpha} \end{cases}.$$

13.2 Критерий Мизеса-Смирнова

Определим тестовую статистику:

$$L(\bar{X}) = \sqrt{n}\rho_2(F_0, F_n).$$

По th. Мизеса-Смирнова:

$$P(L(\bar{X}) < z) \xrightarrow{n \rightarrow \infty} \mathcal{S}(z),$$

где \mathcal{S} – распределение Мизеса-Смирнова. Тогда порогом принятия решения при уровне значимости α является квантиль распределения Мизеса-Смирнова порядка $1 - \alpha$ (далее $s_{1-\alpha}$).

Таким образом, определим тест:

$$\varphi(\bar{X}) = \begin{cases} 0, & \sqrt{n}\rho_2^2(F_0, F_n) < w_{1-\alpha} \\ 1, & \sqrt{n}\rho_2^2(F_0, F_n) \geq w_{1-\alpha} \end{cases}.$$

13.3 Прикладной алгоритм

1. Строится ЭФР
2. Считается статистика критерия. Поскольку ЭФР является кусочно-постоянной, расстояние Колмогорова / Мизеса-Смирнова можно считать как верхнюю границу по соответствующим значениям расстояний в точках скачка.
3. Для заданного уровня значимости α находится квантиль распределения Колмогорова / Мизеса-Смирнова порядка $1 - \alpha$.
4. Если значение тестовой статистики меньше полученного квантиля, следует принять нулевую гипотезу, иначе – отклонить.

14 Выборочный метод построения оценок одномерных характеристик. Асимптотическая нормальность оценки. Построение асимптотического доверительного интервала на базе асимптотической нормальности.

Есть три метода построения оценок:

1. Выборочный метод
2. Метод моментов
3. Метод максимального правдоподобия

14.1 Описание выборочного метода

Этот метод основывается на знании того, что ЭФР $F_n(t)$ является “хорошей” оценкой функции распределения $F_x(t)$.

ЭФР $F_n(t)$ является функцией распределения дискретной случайной величины Y , имеющей следующий ряд распределения:

$$\begin{array}{c|c|c|c|c} Y_i & x_{(1)} & x_{(2)} & \dots & x_{(n)} \\ \hline p_i & 1/n & 1/n & \dots & 1/n \end{array}$$

где $x_{(1)}, \dots, x_{(n)}$ упорядоченная выборка:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

В основе выборочного метода лежит идея: любую характеристику генеральной совокупности X оценивать при помощи соответствующей характеристики случайной величины Y . Естественно полученные таким образом оценки нужно изучать, проверить их свойства. С точки зрения квадратического риска они не всегда являются лучшими в соответствующем классе распределений.

14.2 Асимптотическая нормальность, свойства асимптотической нормальности оценок

Определение. Последовательность оценок \hat{g}_n характеристики $g(\theta)$ называется *асимптотически нормальной* с асимптотической дисперсией $\Delta^2(\theta) > 0$, если случайная величина $Y_n = \sqrt{n}(\hat{g}_n - g(\theta))$ сходится по $P_{n,\theta}$ - распределению к нормальной случайной величине Y с нулевым средним и дисперсией $\Delta^2(\theta)$:

$$(1) Y_n \xrightarrow[n \rightarrow \infty]{P_{n,x}} Y \sim N(0, \Delta^2(p)).$$

Это соотношение можно переписать в виде:

$$\hat{g}_n = g(\theta) + \frac{Y_n}{\sqrt{n}}, \quad Y_n \xrightarrow[n \rightarrow \infty]{P_{n,\theta}} Y \sim N(0, \Delta^2(\theta)),$$

то есть $\hat{g}_n - g(\theta)$ - отклонение оценки от неизвестного значения оцениваемой характеристики имеет приближенно нормальное распределение с нулевым средним и дисперсией $\frac{\Delta^2(\theta)}{n}$.

Определение. Величина $\Delta(\theta) > 0$ называется нормирующим множителем.

$P_{n,\theta}(\frac{Y_n}{\Delta(\theta)} < t) \xrightarrow[n \rightarrow \infty]{} F_{N(0,1)}(t) \Rightarrow P_{n,x}(|\hat{g}_n - g(\theta)| < \frac{T(\Delta(\theta))}{\sqrt{n}}) \xrightarrow[n \rightarrow \infty]{} 2\Phi(T) - 1 = \gamma \Rightarrow T = \frac{1+\gamma}{2}$ - квантиль $N(0, 1)$; γ - надежность, $\delta_n = T_{\frac{1+\gamma}{2}} \frac{\Delta}{(\theta)}$ - точность оценки. $(\hat{g}_n - \delta_n, \hat{g}_n + \delta_n)$ - асимптотически доверительный интервал надежности γ . Если при этом $E_{n,x}(Y_n^2) \xrightarrow[n \rightarrow \infty]{E} (Y^2) = \Delta^2(\theta)(E_{n,x}(Y_n) \rightarrow E(Y) = 0)$, то $E_{n,x}(Y_n) = \sqrt{n}(E_{n,x}(\hat{g})_n - g(\theta)) = \sqrt{n}b_{n,x}(\hat{g})_n \xrightarrow[n \rightarrow \infty]{} 0$ т.е. смещение стремится к нулю быстрее, чем $\frac{1}{\sqrt{n}}$ - \sqrt{n} несмещ \hat{g}_n

$$D_{n,x}(Y_n) = E_{n,x}(Y_n^2) - (E_{n,x}(Y_n))^2 \xrightarrow[n \rightarrow \infty]{} \Delta^2(\theta).$$

$$D_{n,x}(\sqrt{n}(\hat{g}_n - g(\theta))) = nD_{n,x}(\hat{g}_n).$$

- инвариантна относительно сдвига и вынесение \sqrt{n} из-под знака дисперсии т.е. $nD_{n,x}(\hat{g}_n) \rightarrow \Delta^2(\theta)$ $\sqrt{n}\sigma_{n,x}(\hat{g}_n) \rightarrow \Delta(\theta)$, средне квадратическое отклонение имеет порядок $\frac{\Delta(\theta)}{\sqrt{n}}$

$$nR_2(\hat{g}_n, \theta) = n(b_{n,x}^2(\hat{g}_n) + D_{n,x}(\hat{g}_n)) \rightarrow 0 + \Delta^2(\theta).$$

т.е. для асимптотически нормальных оценок дисперсия и квадратический риск в асимптотике совпадают.

3. Без док-ва. утверждение Оценки S_n^2 и σ_n^2 - асимптотически нормальны с $\Delta^2(X) = E(X - EX)^4 - D^2(X)$, если существуют четвертые центральные моменты. т.е. $\frac{\sigma_n^2 - D(X)}{\Delta(X)} \sqrt{n} \xrightarrow[n \rightarrow \infty]{P_{n,x}} Y \sim N(0, 1)$ и точность выборочной дисперсии оценивается моментами 4-го порядка.

15 Основные выборочные оценки и их свойства. Выборочное математическое ожидание. Выборочная дисперсия. Выборочные моменты. Выборочные медиана и квантили. Выборочные оценки ковариации и коэффициента корреляции.

Здесь X – произвольная рассматриваемая случайная величина.

15.1 Выборочное среднее / М.О.

Определение. Случайную величину $\bar{X}_n = EY = \sum_{i=1}^n X_{(i)} \cdot \frac{1}{n} = \frac{1}{n} \sum_{j=1}^n X_j$ называют *выборочным средним* некоторой выборки $X_{[n]}$ из генеральной совокупности X .

Выборочное среднее является выборочной точечной оценкой EX .

Свойства.

- Выборка является набором одинаково распределенных независимых случайных величин, из чего по закону больших чисел:

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} EX, \text{ если } \exists EX.$$

Поэтому выборочное среднее является *состоятельной* оценкой EX .

- Выборочное среднее является *несмещенной* оценкой EX :

$$E_x \bar{X}_n = \frac{1}{n} \sum_{i=1}^n E(X_i) = EX, \text{ оо св-ву М.О.}$$

- Если $\exists EX, DX$, то по центральной предельной теореме:

$$Y_n = \frac{\bar{X}_n - E_x(\bar{X}_n)}{\sigma_x(\bar{X}_n)} = \frac{\bar{X}_n - EX}{\sigma(X)} \sqrt{n} \xrightarrow[n \rightarrow \infty]{F} Y \sim N(0, 1),$$

или

$$F_{Y_n}(t) \xrightarrow[n \rightarrow \infty]{\mathbb{R}} F_Y(t).$$

Из этого следует, что центрированное нормированное выборочное среднее сходится по распределению к стандартному нормальному распределению. Следовательно, выборочное среднее является *асимптотически нормальной* оценкой EX .

15.2 Выборочная дисперсия

Определение. Случайную величину $S_n^2 = D(X_{[n]}) = \sum_{i=1}^n (X_{(i)} - EY)^2 \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ называют *выборочной дисперсией* некоторой выборки $X_{[n]}$ из генеральной совокупности X .

Выборочная дисперсия является выборочной точечной оценкой DX .

Свойства.

- Выборочная дисперсия является *состоятельной* оценкой DX :

$$S_n^2 \xrightarrow[n \rightarrow \infty]{P} E(X^2) - (EX)^2 = DX.$$

- Выборочная дисперсия является *смещенной* оценкой DX .
(Ниже, не теряя общности, будем считать $EX = 0$, инвариантность DX относительно сдвига):

$$E(X^2) = DX, \quad E_x(\bar{X}_n) = EX = 0$$

$$\Rightarrow E_x(\bar{X}_n^2) = D_x(\bar{X}_n) + (E_x(\bar{X}_n))^2 = \frac{DX}{n}$$

$$\Rightarrow E_x S_n^2 = E_x \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right) = \frac{1}{n} n E(X^2) - \frac{DX}{n} = DX - \frac{DX}{n} = \frac{n-1}{n} DX.$$

- Выборочная дисперсия является *асимптотически нормальной* оценкой DX – без доказательства.

15.3 Несмещенная выборочная дисперсия

Определение. Чаще вместо S_n^2 используют *несмещенную (исправленную) оценку дисперсии*:

$$\sigma_n^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

Несмещенная выборочная дисперсия является выборочной точечной оценкой DX .

Свойства.

- *Состоятельность* следует из состоятельности S_n^2 :

$$\sigma_n^2 \xrightarrow[n \rightarrow \infty]{p=1} DX.$$

- *Несмещенность* очевидна из доказательства смещенности выборочного среднего.
- Несмещенная оценка дисперсии является *асимптотически нормальной* оценкой DX – без доказательства.

15.4 Выборочные моменты

15.4.1 Выборочные начальные моменты

Определение. Выборочным начальным моментом порядка k называется статистика:

$$m_{n,k} = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

Эти выборочные характеристики можно считать выборочным средним для случайной величины $Z = X^k$:

$$m_{n,k} = \bar{Z}_k.$$

Следовательно, если $\exists E(X^k)$, то $m_{n,k}$ является состоятельной и несмещенной оценкой $E(X^k)$.

Если существует $E(X^{2k})$, то $m_{n,k}$ является асимптотически нормальной оценкой $E(X^k)$ с асимптотической дисперсией $\Delta^2 = E_x(Z^2) - (E_x(Z))^2$.

15.4.2 Выборочные центральные моменты

Определение. Выборочным центральным моментом порядка k называется статистика:

$$\mu_{n,k} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k.$$

Данные статистики являются состоятельными, смещенными оценками соответствующих центральных моментов генеральной совокупности.

15.5 Выборочная медиана

Определение. Медианой t_0 случайной величины X ($med(x)$) называют такое значение аргумента функции распределения $F_x(t)$, что для него выполняются неравенства:

$$\begin{cases} P(X \geq t_0) \geq \frac{1}{2} \\ P(X \leq t_0) \geq \frac{1}{2} \end{cases}.$$

Если $F_x(t) \in C(\mathbb{R})$, то $F_x(t_0) = \frac{1}{2}$.

Определение. Пусть $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ — упорядоченная выборка (вариационный ряд), тогда выборочной медианой med_n называется следующая случайная величина:

$$med_n = \begin{cases} X_{(k)} = X_{\frac{n-1}{2}}, & \text{при } n = 2k - 1 \\ \frac{X_{(k)} + X_{(k+1)}}{2}, & \text{при } n = 2k \end{cases}.$$

Свойства. Пусть генеральная совокупность является непрерывной случайной величиной и $T = \{t : 0 < F_X(t) < 1\}$. Если $f_X(t)$ непрерывна и положительна при $t \in T$, то плотность распределения случайной величины $Y = f_Y(t)$, где

$$Y = 2\sqrt{n}f_X(t_0)(\text{med}_n - t_0), \quad t_0 = \text{med}(X)$$

при $n \rightarrow \infty$ стремится к $f_{N(0,1)}(t) = \frac{1}{\sqrt{2\pi}} \exp -\frac{t^2}{2}$, а

$$P(a < Y < b) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

- Следовательно, выборочная медиана является *состоятельной* оценкой $\text{med}(X)$.
- Также видно, что выборочная медиана является *асимптотически нормальной* оценкой $\text{med}(x)$ с асимптотической дисперсией $\Delta^2 = \frac{1}{rf_X^2(t_0)}$.

- Выборочная медиана является \sqrt{n} -несмещенной оценкой $\text{med}(X)$. То есть:

$$\sqrt{n}b_{n,\theta}(\text{med}_n) = \sqrt{n}(E_x(\text{med}_n) - \text{med}(X)) \xrightarrow{n \rightarrow \infty} 0.$$

15.6 Выборочная ковариация и корреляция

Выборочная ковариация и корреляция используются при решении вопроса о наличии зависимости между случайными величинами X и Y .

В этом случае рассматривается выборка из случайного вектора (X, Y) . Здесь пары $\{X_i, Y_i\}_i$ независимы и одинаково распределены. Если случайные величины X и Y не являются линейно зависимыми ($r(X, Y) \neq 1$), то для последовательности $\{X_i, Y_i\}_i$ справедливо утверждение аналогичное центральной предельной теореме.

15.6.1 Выборочная ковариация

Определение. Выборочной ковариацией называется статистика:

$$K_n = K_n(X, Y) = \frac{1}{n} \sum_{i=1}^n ((X_i - \bar{X}_n)(Y_i - \bar{Y}_n)) = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X}_n \bar{Y}_n.$$

Свойства.

- *Состоятельная* оценка. X и Y .
- *Смещенная* оценка. Аналогично выборочной дисперсии, можно показать:

$$E_x(K_n) = \frac{n-1}{n} K(X, Y).$$

- *Асимптотически нормальная* оценка.

Определение. В приложениях обычно рассматривают *несмещенную* оценку ковариации:

$$\tilde{K}_n = \frac{n}{n-1} K_n = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \bar{X}_n)(Y_i - \bar{Y}_n)).$$

15.7 Выборочная корреляция

Определение. Выборочной корреляцией X и Y называется статистика:

$$r_n = r_n(X, Y) = \frac{K_n}{S_{n,X}S_{n,Y}} = \frac{\tilde{K}_n}{\sigma_{n,X}\sigma_{n,Y}}.$$

Замечание. В определении выше предполагается существование всех необходимых моментов: $EX, EY, K(X, Y), \dots$

Свойства.

- *Состоятельная оценка.*
- *Несмещенная оценка.*
- *Асимптотически нормальная оценка.*

16 Гистограмма как оценка плотности распределения. Статистические свойства гистограммы. Теорема Пирсона. Критерий хи-квадрат для проверки гипотезы о виде распределения генеральной совокупности

16.1 Построение

Основной идеей, использующейся в этом методе, является идея *группировки данных*. Пусть распределение абсолютно непрерывно с непрерывной плотностью распределения $f(x)$. Тогда значение плотности распределения в точке t можно оценить как отношение вероятности попадания значения в полуинтервал $\Delta = [t_1, t_2) \ni t$ к длине этого полуинтервала $t_2 - t_1 = |\Delta|$. Иными словами:

$$f(t) \approx \frac{P(\Delta)}{|\Delta|}.$$

Это приближение можно объяснить следующим образом, пользуясь теоремой Лагранжа:

$$P(\Delta) = \int_{t_1}^{t_2} f(x) dx = f(t_1 + \theta|\Delta|)|\Delta| \approx f(t)|\Delta|.$$

Построим наконец оценку, взяв в качестве $P(x < t_i) = F(t_i)$ выборочное значение:

$$P(\Delta) = F(t_2) - F(t_1) \approx F_n(t_2) - F_n(t_1) = k(\Delta).$$

За $k(\Delta)$ обозначим число элементов выборки, попавших в отрезок Δ .

Определение. *Интервалами группировки* называется разбиение $\{\Delta_0, \Delta_{\pm 1}, \Delta_{\pm 2}, \dots\}$ отрезка $[a, b]$ на дизъюнктные интервалы фиксированной длины $h > 0$.

Определение. *Гистограммой* называется функция $f_n(t)$, принимающая постоянные значения на заданных интервалах группировки:

$$t \in \Delta_m \implies f_n(t) = f_{n,m} = \frac{k(\Delta_m)}{nh}.$$

Замечание. Гистограмма – кусочно постоянная функция.

Теорема 16.1. Гистограмма является плотностью распределения.

Доказательство. $f_n(t) \geq 0$,

$$\int_{\mathbb{R}} f_n(t) dt = \sum_m \int_{\Delta_m} f_n(t) dt = \sum_m h f_{n,m} = n^{-1} \sum_m k(\Delta_m) = 1.$$

■

Замечание. На практике удобно выбирать границы $[a, b]$ в виде максимума и минимума элементов выборки.

16.2 Статистические свойства гистограммы

Гистограмма является оценкой плотности распределения. Изучим её свойства как оценки. Для этого изучим квадратичное отклонение $R_{n,2}(t)$. В нашем случае $g(\theta) = f(t)$. Ранее было показано, что в случае одномерной оценки квадратичный риск представим в виде

$$R_{n,2}(t) = D_n(t) + b_n^2(t), \quad D_n(t) = D_F(f_n(t)), \quad b_n(t) = E_F(f_n(t)) - f(t).$$

Заметим, что при фиксированном t $k(\Delta_m)$ – случайная величина, имеющая биномиальное распределение $k(\Delta_m) \sim B(n, p)$, $p = p_{n,m} = P_F(\Delta_m)$. Отсюда имеем:

$$E_F(k(\Delta_m)) = np, \quad D_F(k(\Delta_m)) = np(1-p).$$

Вычислим на основе этих знаний значения сдвига и дисперсии:

$$b_n(t) = \left(\frac{p}{h} - f(t) \right), \quad D_n(t) = \frac{p(1-p)}{nh^2} \leq \frac{p}{h} \frac{1}{nh}.$$

Имея непрерывность $f(x)$ на отрезке Δ_m по теореме Лагранжа имеем

$$\frac{p}{h} = \frac{1}{h_n} \int_{\Delta_m} f(x) dx = f(\tilde{t}), \quad \tilde{t} \in \Delta_m.$$

Отсюда при условиях $h = h_n \rightarrow 0$, $nh_n \rightarrow +\infty$ следует:

$$b_n(t) = f(\tilde{t}) - f(t) \xrightarrow{n \rightarrow +\infty} 0$$

$$D_n(t) = \frac{f(\tilde{t})}{nh} \xrightarrow{n \rightarrow +\infty} 0.$$

Отсюда вытекает:

$$R_{n,2}(t) \xrightarrow{n \rightarrow +\infty} 0.$$

В этом случае по теореме о достаточном условии состоятельности оценки следует

Теорема 16.2. (Состоятельность гистограммы как оценки f)

Пусть задано абсолютно непрерывное распределение с плотностью $f(x)$, отрезок $[a, b]$ и его разбиение с длинами интервалов h_n такими, чтобы выполнялись условия:

$$h_n \xrightarrow{n \rightarrow +\infty} 0, \quad nh_n \xrightarrow{n \rightarrow +\infty} +\infty.$$

Тогда соответствующая гистограмма является состоятельной оценкой плотности распределения.

Теорема 16.3. Наилучшая скорость убывания длины интервалов группировки в классе плотностей с условием

$$\exists C: \int_{\mathbb{R}} (f'(t))^2 dt \leq C^2$$

имеет порядок $n^{-1/3}$.

16.3 Критерий хи-квадрат

16.3.1 Дискретная случайная величина

Пусть генеральная совокупность X – дискретная случайная величина с распределением $P_X(t = t_k) = p_k$, где \bar{p} – набор неизвестных вероятностей. Пусть решается вопрос о справедливости гипотезы $p = \bar{p}_0 = (p_{0,1}, p_{0,2}, \dots, p_{0,k})$, $p_{0,j} > 0$. Через \mathbb{P} обозначим множество:

$$\mathbb{P} = \{p \in \mathbb{R}^k \mid p_j \geq 0, \sum_j p_j = 1\}.$$

Поставим задачу проверки согласия с $H_0 \equiv \bar{p} = \bar{p}_0$. Пусть n_j – число элементов выборки $X^{(n)}$, принимающих значение t_j , $F_0(t)$ – функция распределения генеральной совокупности при условии H_0 .

Определение. Статистикой хи-квадрат с $k-1$ степенью свободы называется статистика

$$\chi_{n,k-1}^2(X^{(n)}) = \sum_{j=1}^k \frac{(n_j - np_{0,j})^2}{np_{0,j}}.$$

Определение. Функцией распределения хи-квадрат с $k-1$ степенью свободы χ_{k-1}^2 называется функция распределения случайной величины

$$\tau_k = \sum_{i=1}^k \zeta_i^2, \quad \zeta_i \sim N(0, 1).$$

Теорема 16.4. (Пирсон)

Пусть справедливо $\bar{p} = \bar{p}_0$. Тогда справедливо

$$\sup_{u \in \mathbb{R}_{>0}} \left| P_{F_0}(\chi_{n,k-1}^2 < u) - \chi_{k-1}^2(u) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Определение. Критерием хи-квадрат асимптотического уровня значимости α для проверки согласия с гипотезой $H_0 \equiv \bar{p} = \bar{p}_0$ называется последовательность тестов

$$\psi_n(X^{(n)}) = \begin{cases} 1, & \chi_{n,k-1}^2 \geq t_{k-1,\alpha} \\ 0, & \chi_{n,k-1}^2 < t_{k-1,\alpha} \end{cases}.$$

Здесь величина $t_{k-1,\alpha}$ определяется из условия

$$\chi_{k-1}^2(t_{k-1,\alpha}) = 1 - \alpha.$$

Теорема 16.5. (Состоятельность критерия хи-квадрат)

Критерий хи-квадрат является состоятельным критерием асимптотического уровня значимости α .

Доказательство.

- Оценим вероятность ошибки первого рода.

$$\alpha(\psi_n) = P_{n,F_0}(\chi_{n,k-1}^2 \geq t_{k-1,\alpha}) = 1 - P_{n,F_0}(\chi_{n,k-1}^2 < t_{k-1,\alpha}) \xrightarrow{n \rightarrow +\infty} 1 - \chi_{k-1}^2(t_{k-1,\alpha}) = \alpha.$$

Таким образом, критерий имеет асимптотический уровень значимости α .

- Оценим ошибку второго рода. Зафиксируем альтернативу $H_1 \equiv \bar{p} = \bar{p}_1 \neq \bar{p}_0$. Пусть $j_0: p_{1,j_0} \neq p_{0,j_0}, |p_{1,j_0} - p_{0,j_0}| = a$. В силу закона больших чисел $n_{j_0}/n \rightarrow p_{1,j_0}$ почти везде по мере P_F . Поэтому верно

$$(n_{j_0} - np_{0,j_0})^2 \sim n^2 a^2.$$

Откуда по определению следует

$$\chi_{n,k-1}^2 \xrightarrow{n \rightarrow +\infty} +\infty.$$

Поэтому:

$$\beta(\psi_n, F) = P_{n,F}(\chi_{n,k-1}^2 < t_{k-1,\alpha}) \xrightarrow{n \rightarrow +\infty} 0.$$

Таким образом, критерий является состоятельным. ■

16.3.2 Критерий хи-квадрат для случайной величины общего вида

Рассмотрим теперь случайную величину общего вида. Пусть основная гипотеза является простой и имеет вид $H_0 \equiv F_X(x) = F_0(x)$. Чтобы применить критерий хи-квадрат к такой задаче, используют *дискретизацию данных*. Множество значений X разбивается на k множеств, попадание в каждое из которых интерпретируется как значение дискретной случайной величины с k значениями. Для этой случайной величины мы уже умеем применять критерий хи-квадрат.

17 Метод моментов и его свойства.

17.1 Идея метода подстановки

Метод подстановки уже использовался нами в следующих задачах:

- Оценка характеристик распределения $g(F)$ через характеристики выборочного распределения $g(F_n)$.
- Если $\hat{\theta}_n$ – в определенном смысле хорошая оценка параметра распределения θ , мы используем в качестве оценки $g(\theta)$ значение $g(\hat{\theta}_n)$. (*подставляем вместо θ $\hat{\theta}$*).

К этому методу можно подойти и с другой стороны.

17.2 Метод моментов

Пусть мы ищем параметр распределения θ , причем его можно задать как решение уравнения

$$E_{\theta}(H(X, \theta)) = 0.$$

Здесь $H: \mathbb{R} \rightarrow \mathbb{R}$ – известная нам функция. Метод состоит в том, чтобы *заменить* математическое ожидание его выборочной оценкой, то есть в качестве оценки параметра $\hat{\theta}(X^{(n)})$ взять решение уравнения

$$\sum_{i=1}^n H(X_i, \theta) = 0.$$

Сформулируем эти идеи в более общем виде.

Пусть распределение генеральной совокупности F_X известно нам с точностью до неизвестного параметра $\theta \in \Theta \subseteq \mathbb{R}^m$. Понятно, что все числовые характеристики распределения $g(F_X)$ можно выразить через неизвестный нам параметр θ : $g(F_X) = g(\theta)$. Пусть выбранная нами характеристика $g: \mathbb{R}^m \rightarrow \mathbb{R}^k$ удовлетворяет следующим свойствам:

- Система уравнений относительно θ :

$$g_i(\theta) = g_i^0, \quad i = 1..k,$$

где $g^0 \in \mathbb{R}^k$ – теоретическое значение характеристики, имеет единственное решение.

- Система уравнений обладает свойством *устойчивости*, то есть отображение, ставящее в соответствие g^0 решение непрерывно в окрестности g^0 .

В таком случае, заменим g^0 его выборочным аналогом \hat{g}_n^0 . Решим ту же самую систему уравнений:

$$g_i(\theta) = \hat{g}_n^0, \quad i = 1..k.$$

Остается просто взять в качестве оценки неизвестного параметра θ найденное нами решение $\hat{\theta}_n$.

Теорема 17.1. (Свойства метода моментов)

Из определения метода моментов сразу вытекают его основные свойства.

- Если \hat{g}_n – состоятельные оценки, то $\hat{\theta}$ – состоятельная оценка.
- Аналогичное утверждение справедливо и для свойства асимптотической нормальности.

Доказательство.

- Это свойство – непосредственное следствие устойчивости системы.
- Асимптотическая нормальность \hat{g}_n означает

$$\hat{g}_n = g(\theta) + n^{-\frac{1}{2}}Y_n, \quad Y_n \xrightarrow[n \rightarrow +\infty]{P_{n,\theta}} Y \sim N(0, \mathcal{K}(\theta)).$$

Здесь $\mathcal{K}(\theta)$ – матрица ковариаций. Чтобы доказать утверждение, нам достаточно представить оценку в виде

$$\hat{\theta}_n = \theta + n^{-\frac{1}{2}}Z_n.$$

Где $Z_n \sim N(0, _)$. По формуле Тейлора:

$$g\left(\theta + n^{-\frac{1}{2}}Z_n\right) = g(\theta) + g'(\theta)n^{-\frac{1}{2}}Z_n + \mathcal{O}(n^{-1}).$$

С другой стороны, поскольку $\hat{\theta}_n$ является решением соответствующей системы уравнений:

$$g(\hat{\theta}) = \hat{g}_n = g(\theta) + n^{-\frac{1}{2}}Y_n.$$

Приравнивая правые части последних двух уравнений, получаем

$$Z_n \approx (g'(\theta))^{-1}Y_n \xrightarrow[n \rightarrow +\infty]{P_{n,\theta}} Z \sim N(0, R(\theta)).$$

Где

$$R(\theta) = (g'(\theta))^{-1}\mathcal{K}(\theta)(g'(\theta)^\top)^{-1}.$$

■

18 Метод максимального правдоподобия и его свойства.

Будем основывать метод на *принципе максимального правдоподобия*: в качестве оценки неизвестного параметра распределения выберем то значение, при котором вероятность наблюдаемых величин наиболее вероятна.

Будем считать, что выполнено одно из двух:

- Распределение генеральной совокупности абсолютно непрерывно, то есть существует непрерывная плотность, задающая это распределение:

$$f(x, \theta) \Longleftrightarrow P_\theta.$$

- Распределение дискретно. В таком случае будем обозначать

$$f(x, \theta) = P_\theta(X = x).$$

Определение. *Функцией правдоподобия* называется функция

$$L(\theta, X) = f(X, \theta).$$

Определение. *Логарифмической функцией правдоподобия* называется функция

$$l(\theta, X) = \ln L(\theta, X) = \ln f(X, \theta).$$

Замечание. При фиксированном $X \in \mathcal{X}$ функции правдоподобия – просто вещественные функции θ . Если же считать X случайной величиной, то и функции правдоподобия становятся случайными величинами.

Замечание. В модели независимой однородной выборки функции правдоподобия принимают вид:

$$L(\theta, X^{(n)}) = \prod_{i=1}^n f(X^{(n)}, \theta), \quad l(\theta, X^{(n)}) = \sum_{i=1}^n \ln f(X^{(n)}, \theta).$$

Определение. *Оценкой максимального правдоподобия* называется значение

$$\theta^*(X) = \operatorname{argmax}_{\theta \in \Theta} L(\theta, X).$$

Определение. В случае, когда логарифмическая функция правдоподобия непрерывно дифференцируема, система уравнений

$$\frac{\partial l(\theta, X)}{\partial \theta_j} = 0$$

Называется *уравнениями максимального правдоподобия*. В этом случае $\theta^*(X)$ является одним из решений этой системы.

Определение. Информацией Фишера называется функция

$$I(\theta) = E_{\theta}(l'(\theta, X))^2.$$

Замечание. Информация Фишера – числовая характеристика распределения, и не является случайной величиной.

Определение. Оценка называется $\alpha(n)$ -несмещенной, если

$$\alpha(n)b_{n,\theta}(\hat{g}_n) \xrightarrow{n \rightarrow +\infty} 0.$$

Теорема 18.1. (Свойства оценки максимального правдоподобия)

Пусть справедливы условия:

- $\theta \in \Theta = \langle a, b \rangle \subseteq \mathbb{R}$, то есть изучаемый параметр одномерный.
- $\mathcal{X} = \mathbb{R}$.
- Почти везде существуют частные производные логарифмической функции правдоподобия порядка $k \leq 3$.
- Выполнены неравенства

$$\left| \frac{\partial^k l(\theta, X)}{\partial \theta^k} \right| \leq G_k(x), \quad 1 \leq k \leq 3.$$

Причем G_k суммируемы и

$$\sup_{\theta \in \Theta} \int_{\mathbb{R}} G_3(x) f(x, \theta) dx < +\infty.$$

- $\forall \theta > 0 \exists I(\theta) > 0$.

Тогда соответствующая оценка максимального правдоподобия обладает свойствами:

- Состоятельность.
- \sqrt{n} -несмещенность.
- Асимптотическая нормальность с $\Delta^2(\theta) = \frac{1}{I(\theta)}$.

19 О сравнении качества оценок. Свойства функции правдоподобия (одномерный параметр). Неравенство Рао-Крамера и эффективные оценки.

Рассматриваются задачи оценки конечномерного параметра распределения P_θ , $\theta \in \Theta \subset R^m$, а также характеристик (функций) $g(\theta)$ по наблюдениям $X \in \mathcal{X}$.

19.1 О сравнении качества оценок

Сравниваем различные оценки с помощью функции риска.

Определение. Оценка \hat{g}^1 не хуже оценки \hat{g}^2 , если $R(\hat{g}^1, \theta) \leq R(\hat{g}^2, \theta)$, для всех $\theta \in \Theta$. Обозначение: $\hat{g}^1 \succeq \hat{g}^2$.

Определение. Пусть $G = \{\hat{g}\}$ – некоторый класс оценок. Оценка \hat{g}^* называется *эффективной* в классе G , если:

$$\hat{g}^* \succeq \hat{g}$$

для всех $\hat{g} \in G$.

В классе *всех* оценок не существует эффективной оценки. Для поиска эффективных оценок нужны ограничения на класс рассматриваемых оценок.

Отмеченные трудности вынуждают сравнивать не функции риска различных оценок, а какие-нибудь числовые величины от функций риска, которые характеризуют функцию риска.

19.1.1 Минимаксный подход

Здесь качество оценки характеризуется максимальным значением риска:

$$R_{\max}(\hat{g}) = \sup_{\theta \in \Theta} R(\hat{g}, \theta).$$

Определение. Оценка \hat{g} называется *минимаксной*, если:

$$R_{\max}(\hat{g}) \leq R_{\max}(\tilde{g})$$

для любой оценки \tilde{g} .

Подход ориентирован на построение оценки с минимальным значением максимального риска.

19.1.2 Асимптотически минимаксные оценки

Определение. Оценка \hat{g}_n называется *асимптотически минимаксной*, если:

$$\lim_{n \rightarrow +\infty} \left(\frac{R_{\max}(\hat{g}_n)}{R_{\max}(\tilde{g}_n)} \right) \leq 1$$

для любой оценки \tilde{g} .

Определение. Оценка \hat{g}_n называется *локально асимптотически минимаксной* в точке $\theta_0 \in \Theta$, если для достаточно малых $\varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} \left(\frac{\sup_{|\theta - \theta_0| \leq \varepsilon} R(\hat{g}_n, \theta)}{\sup_{|\theta - \theta_0| \leq \varepsilon} R(\tilde{g}_n, \theta)} \right) \leq 1$$

для любой оценки \tilde{g} .

19.2 Свойства функции правдоподобия (одномерный параметр)

TODO

19.3 Неравенство Рао-Крамера и эффективные оценки

Пусть $\hat{\theta} = \hat{\theta}(X)$ – несмещенная оценка одномерного параметра, выполнены условия регулярности и $I(\theta) > 0$ для всех $\theta \in \Theta$.

Теорема 19.1. (Неравенство Рао-Крамера)

Для любого $\theta \in \Theta \subset \mathbb{R}$

$$D_\theta(\hat{\theta}) \geq \frac{1}{I(\theta)}.$$

Доказательство.

- Будем считать, что наблюдения X имеют непрерывное распределение. Запишем условие несмещенности:

$$\int_{\mathbb{X}} \hat{\theta}(x) f(x, \theta) dx = \theta$$

- Продифференцируем это равенство:

$$\int_{\mathbb{X}} \hat{\theta}(x) l'(\theta, x) f(x, \theta) dx = 1$$

- Умножая 19.2.3 на θ и вычитая из п.2:

$$\int_{\mathbb{X}} (\hat{\theta}(x) - \theta) l'(\theta, x) f(x, \theta) dx = 1$$

- Рассмотрим функции:

$$g_1(x) = (\hat{\theta}(x) - \theta)(f(x, \theta))^{\frac{1}{2}}, \quad g_2(x) = l'(x, \theta)(f(x, \theta))^{\frac{1}{2}}$$

- Возведем в квадрат п.3 и воспользуемся интегральным неравенством Коши-Буняковского для функций п.4:

$$1 \leq \int_{\mathbb{X}} (\hat{\theta}(x) - \theta)^2 f(x, \theta) dx \int_{\mathbb{X}} (l'(\theta, x))^2 f(x, \theta) dx = D_\theta(\hat{\theta}) I(\theta).$$



Неравенство Рао-Крамера дает нижнюю границу для дисперсии и квадратичного риска несмещенных оценок.

Определение. Оценка, на которой достигается нижняя граница Рао-Крамера, называется *эффективной*.

Замечание. Неравенство Рао-Крамера справедливо и для смещенных оценок (со смещением $b'_\theta(\hat{\theta})$) в форме:

$$D_\theta(\hat{\theta}) \geq \frac{(1 + b'_\theta(\hat{\theta}))^2}{I(\theta)}.$$

20 Примеры построения наиболее мощных и равномерно наиболее мощных тестов.

20.1 Пример 1

Имеется выборка X_1, \dots, X_n из нормального распределения со средним 0 и дисперсией σ^2 , $\sigma > 0$. Построим наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \sigma = \sigma_1$ против альтернативы $H_2 = \sigma = \sigma_2$, где $\sigma_1 < \sigma_2$.

Отношение правдоподобия имеет абсолютно непрерывное распределение при любой из гипотез, поэтому критерий отношения правдоподобия будет нерандомизированным. Его критическая область $\delta(X) = H_2$ определяется неравенством:

$$T(X) = \frac{\sigma_1^n}{\sigma_2^n} \exp\left(\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\sum_{i=1}^n X_i^2\right),$$

что эквивалентно неравенству $\overline{X^2} \geq c_1$. Найдем c_1 , при котором размер критерия равен ε :

$$\alpha_1(\delta) = P_{H_1}(\overline{X^2} \geq c_1) = P_{H_1}\left(\frac{n\overline{X^2}}{\sigma_1^2} \geq \frac{nc_1}{\sigma_1^2}\right) = 1 - H_n\left(\frac{nc_1}{\sigma_1^2}\right) = \varepsilon.$$

Отсюда $n\frac{c_1}{\sigma_1^2} = h_{1-\varepsilon}$ – квантиль χ^2 -распределения с n степенями свободы уровня $1 - \varepsilon$. Тогда $c_1 = \frac{h_{1-\varepsilon}\sigma_1^2}{n}$ и НМК размера ε имеет вид $\delta(X) = H_2$ при:

$$\overline{X^2} > \frac{h_{1-\varepsilon}\sigma_1^2}{n}.$$

21 Доверительное оценивание и проверка гипотез на основе оценок максимального правдоподобия.

21.1 Доверительное оценивание

Напомним утверждение, сформулированное ранее.

Теорема 21.1. При достаточно общих условиях оценка максимального правдоподобия одномерного параметра обладает свойством асимптотической нормальности с $\Delta^2(\theta) = \frac{1}{I(\theta)}$.

Обладая этой информацией, нетрудно построить доверительный интервал на основе оценки \hat{g} , полученной методом максимального правдоподобия. Действительно, мы только что сформулировали тот факт, что

$$\sqrt{n} \cdot (\hat{g} - g(\theta)) = Y_n \xrightarrow[n \rightarrow +\infty]{P_{n,\theta}} Y \sim N(0, \Delta^2(\theta)).$$

Откуда вытекает

$$P_{n,\theta}(|Y_n/\Delta(\theta)| \leq t_\gamma) \rightarrow \gamma.$$

Здесь $t_\gamma: P_{N(0,1)}(|\xi| < t_\gamma) = \gamma$, $\xi \sim N(0, 1)$. Чтобы в явном виде получить доверительный интервал, раскроем определение Y_n и подставим значение $\Delta(\theta)$:

$$\begin{aligned} P_{n,\theta} \left(\left| \sqrt{n} \cdot (\hat{g} - g(\theta)) \cdot \sqrt{I(\theta)} \right| \leq t_\gamma \right) &\rightarrow \gamma \\ \iff P_{n,\theta}(\hat{g} - \delta \leq g(\theta) \leq \hat{g} + \delta) &\rightarrow \gamma, \quad \delta = \frac{t_\gamma}{\sqrt{nI(\theta)}}. \end{aligned}$$

21.2 Проверка гипотез

Пусть сформулирована гипотеза $H_0 \equiv \theta = \theta_0$ и альтернатива $H_1 \equiv \theta = \theta_1 > \theta_0$ (Такая альтернатива называется *правосторонней*). Пусть также имеется оценка $\hat{\theta}$, полученная методом максимального правдоподобия. Рассмотрим тест:

$$\psi_{n,\alpha}^* = \begin{cases} 1, & \sqrt{nI(\theta)} \cdot (\hat{\theta} - \theta_0) \geq c_{1-\alpha} \\ 0, & \sqrt{nI(\theta)} \cdot (\hat{\theta} - \theta_0) < c_{1-\alpha} \end{cases}.$$

Здесь $c_\gamma: P(\xi < c_\gamma) = \gamma$, $\xi \sim N(0, 1)$.

Теорема 21.2. $\psi_{n,\alpha}^*$ является состоятельным тестом асимптотического уровня значимости α .

Доказательство.

- Вычислим уровень значимости критерия:

$$\alpha(\psi_{n,\alpha}^*) = P_{n,\theta_0} \left(\sqrt{nI(\theta)} \cdot (\hat{\theta} - \theta_0) \geq c_{1-\alpha} \right) = 1 - P_{n,\theta_0} \left(\sqrt{nI(\theta)} \cdot (\hat{\theta} - \theta_0) < c_{1-\alpha} \right) \rightarrow \alpha.$$

Последнее верно по свойству асимптотической нормальности оценки.

- Проверим состоятельность теста:

$$\begin{aligned}\beta(\psi_{n,\alpha}^*) &= P_{n,\theta_1}(\sqrt{nI(\theta)} \cdot (\hat{\theta} - \theta_0) < c_{1-\alpha}) \\ &= P_{n,\theta_1}\left(\sqrt{nI(\theta)} \cdot (\hat{\theta} - \theta_1) < \underbrace{\sqrt{nI(\theta)} \cdot (\theta_0 - \theta_1)}_{\rightarrow -\infty} + c_\gamma\right) \xrightarrow{n \rightarrow +\infty} 0.\end{aligned}$$

■

22 Общая линейная модель или задачи регрессии

В приложениях часто возникают задачи о наблюдениях, зависящих от изменяющихся параметров эксперимента. Пусть проводится n экспериментов с m изменяющимися параметрами, причем в i -м из них набор параметров выглядит следующим образом:

$$x_i = (x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}).$$

Тогда всем экспериментам сразу соответствует матрица

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,m} \\ x_{2,1} & x_{2,2} & \dots & x_{2,m} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,m} \end{pmatrix}.$$

Эта матрица не является набором случайных чисел.

Определение. Задачи регрессии или общие линейные модели соответствуют предположению, что наблюдаемые в i -м эксперименте данные y_i зависят от параметров линейно с точностью до нормально распределенной с нулевым матожиданием ошибки $\xi_i \sim N(0, \sigma^2)$, причем ξ_i независимы:

$$y_i = \sum_{k=1}^m \theta_k x_{i,k} + \xi_i.$$

Или в матричном виде сразу для всех n экспериментов:

$$Y = X\theta + \xi.$$

В рамках задачи регрессии требуется по данным X, Y восстановить набор коэффициентов $\theta = (\theta_1, \dots, \theta_m)$.

Определение. В рамках текущей задачи X называется *регрессором*, а Y – *откликом*.

Определение. Можно поставить задачу линейной регрессии по-другому. Пусть имеется m функций $\varphi_k: \mathbb{R} \rightarrow \mathbb{R}$ и одномерный параметр z_i . Тогда нужно восстановить зависимость такого вида:

$$y_i = \sum_{k=1}^m \theta_k \varphi_k(z_i) + \xi_i.$$

Построим решение задачи линейной регрессии со следующими дополнительными условиями:

- $Y_i \sim N(m_i, \sigma^2)$.
- Набор параметров X подобран таким образом, чтобы его столбцы были линейно независимы.

Для оценивания коэффициентов θ_i применим метод максимального правдоподобия. Вычислим функцию правдоподобия:

$$L(Y) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - m_i)^2 \right), \quad m_i = \sum_{k=1}^m \theta_k x_{i,k}.$$

Тогда логарифмическая функция правдоподобия имеет вид

$$l(Y) = \ln L(Y) = C - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(Y_i - \sum_{k=1}^m \theta_k x_{i,k} \right)^2.$$

Для минимизации функции правдоподобия осталось найти наименее удаленную от Y точку в пространстве \mathcal{L} , натянутом на x_1, \dots, x_m . Обозначим искомую точку за $\beta = X\theta^*$. Тогда необходимым условием минимума является:

$$Y - \beta \perp \mathcal{L}.$$

Это условие эквивалентно системе

$$\forall k \quad Y - \beta \perp x_k.$$

Раскрываем ортогональность:

$$\forall k \quad \langle Y - \beta, x_k \rangle = 0.$$

И переписываем в матричном виде:

$$X^\top \cdot (Y - X\theta^*) = 0.$$

Выражаем θ^* :

$$\theta^* = (X^\top X)^{-1} \cdot X^\top Y.$$

Матрица $X^\top X$ положительно определена в случае линейной независимости x_i , поэтому решение существует и единственно.

23 Простейшие случайные процессы. Общие определения. Примеры. Моменты.

Определение. Случайным процессом называется семейство случайных величин, заданных на одном вероятностном пространстве, параметризованное $t \in T$:

$$\mathcal{P} = \{ \xi_t(\omega) \mid t \in T \}.$$

Пример.

- Последовательность независимых случайных величин есть случайный процесс с $T = \mathbb{N}$. Процессы, в которых $T \subseteq \mathbb{Z}$, называются *дискретными*.
- Эмпирическая функция распределения – случайный процесс с $T = \mathbb{R}$.
- Случайная величина – частный случай случайного процесса с $T = \{1\}$.

Определение. При заданном $\omega_0 \in \Omega$ *траекторией* называется функция

$$\xi(t) = \xi_t(\omega_0).$$

Определение. При заданном $t_0 \in T$ *сечением* называется случайная величина

$$\xi(\omega) = \xi_{t_0}(\omega).$$

Пример.

- Пусть $X \sim U(-1, 1)$, тогда рассмотрим процесс

$$Y(t, \omega) = X(\omega)e^{-t}.$$

В ситуации, когда случайная величина входит в качестве параметра, процесс называется *простейшим*, или *элементарным*. Траектории устроены следующим образом:

$$\xi(t) = ae^{-t}, \quad a = X(\omega_0).$$

Сечение выглядят так:

$$\xi(\omega) = X(\omega)a, \quad a = e^{-t_0}.$$

- Пусть $X \sim N(a, \sigma^2)$, $Y(t, \omega) = X(\omega)e^{-t}$. Найдем моменты сечений:
 - $E_Y(t) = E(Xe^{-t}) = e^{-t} \cdot a$.
 - $D_Y(t) = D(Xe^{-t}) = e^{-2t} \cdot \sigma^2$.
 - $\dot{Y}(t) = Y(t) - E_Y(t) = (X - a)e^{-t} = \dot{X}e^{-t}$.
 - $K_Y(t, t') = E(\dot{Y}(t) \cdot \dot{Y}(t')) = e^{-t-t'} \cdot E(\dot{X}^2) = e^{-t-t'} \cdot \sigma^2$. Эта функция называется *корреляционной функцией случайного процесса $Y(t)$* .
 - $r_Y(t, t') = \frac{\sigma^2 e^{-t-t'}}{\sigma_Y(t) \cdot \sigma_Y(t')} = \frac{\sigma^2 e^{-t-t'}}{\sigma e^{-t} \sigma e^{-t'}} = 1$. В данном случае получается, что зависимость между сечениями имеет линейный вид.

Найдем распределение сечения. $Y(t_0, \omega) = X e^{-t_0}$ – нормальная случайная величина:

$$Y(t_0, \omega) \sim N(ae^{-t_0}, \sigma^2 e^{-2t_0}).$$

С соответствующим распределением:

$$f_{Y, t_0}(y) = \frac{1}{\sqrt{2\pi}\sigma e^{-t_0}} \left(\exp - \frac{1}{2} \left(\frac{y - ae^{-t_0}}{\sigma e^{-t_0}} \right)^2 \right).$$

24 Цепи Маркова. Марковская зависимость. Переходные вероятности. Предельные вероятности. Схемы блужданий.

24.1 Марковская зависимость

Пусть G – эксперимент с конечным числом исходов E_i . Будем неограниченно повторять эксперимент G . Рассмотрим связанную с этим процессом последовательность случайных величин X_i , $i \in \mathbb{N}_0$. Будем для краткости обозначать $E_i = i$ там, где это удобно.

Определение. Последовательность случайных величин X_i образует *цепь Маркова*, если выполнено

$$P(X_l = j \mid X_0 = k_0, \dots, X_{l-1} = i) = P(X_l = j \mid X_{l-1} = i) \stackrel{\text{def}}{=} p_{i,j}^{(l)}.$$

Определение. Начальное состояние цепи Маркова задается распределением X_0 :

$$P(X_0 = i) = p_i(0).$$

Замечание. Отличие марковской цепи от других дискретных процессов состоит в том, что результат очередного эксперимента зависит исключительно от результата предыдущего эксперимента.

Определение. Вероятности перейти из одного состояния в другое за один шаг можно представить в виде *матрицы переходов*

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix}.$$

Замечание. Отметим пару очевидных свойств матрицы переходов:

- $p_{i,j} \geq 0$.
- $\forall i \sum_{j=1}^n p_{i,j} = 1$.

Определение. Матрица, удовлетворяющая вышеперечисленным свойствам, называется *стохастической*.

Замечание. Понятно, что любая матрица переходов является стохастической. Кроме того, любая стохастическая матрица является матрицей переходов соответствующей марковской цепи.

Определение. Обозначим $p_{i,j}(k)$ – вероятность перехода из состояния i в состояние j за k шагов:

$$p_{i,j}(k) = P(X_{k+l} = j \mid X_l = i) = P(X_k = j \mid X_0 = i).$$

Матрицу из таких вероятностей обозначим P_k .

Теорема 24.1. $P_k = P^k$.

Доказательство. По формуле полной вероятности при $k > 1$ имеем

$$p_{i,j}(k+1) = \sum_{l=1}^n P(X_k = l \mid X_0 = i) p_{l,j} = \sum_{l=1}^n p_{i,l}(k) p_{l,j}.$$

В матричном виде

$$P_{k+1} = P_k \cdot P.$$

Откуда по индукции с базой $P_1 = P$ получаем требуемое. ■

Замечание. Если задано начальное распределение $p_i(0) = P(X_0 = i)$, то аналогичным образом можно получить распределение вероятностей для произвольного момента t :

$$p_i(t) = P(X_t = i) = \sum_{l=1}^n p_l(0) p_{l,i}(t).$$

24.2 Предельные вероятности

Определение. Цепь Маркова, для которой для всех j существует предел

$$p_j = \lim_{t \rightarrow +\infty} p_{i,j}(t).$$

называется *эргодической*.

Теорема 24.2. Если цепь Маркова является эргодической, то система уравнений

$$x_j = \sum_{l=1}^n x_l p_{l,j}, \quad \sum_{l=1}^n x_l = 1$$

имеет единственное решение

$$x_j = p_j = \lim_{t \rightarrow +\infty} p_{i,j}(t).$$

Доказательство.

- Покажем, что $x_j = p_j$ действительно решение системы. Для этого выпишем равенства из доказательства теоремы $P_k = P^k$:

$$p_{i,j}(k+1) = \sum_{l=1}^n p_{i,l}(k) p_{l,j}.$$

И перейдем в нём к предельному переходу при $k \rightarrow +\infty$:

$$p_j = \sum_{l=1}^n p_l p_{l,j}.$$

Для проверки последнего уравнения системы перейдем к пределу при $k \rightarrow +\infty$ в уравнении

$$\sum_{j=1}^n p_{i,j}(k) = 1.$$

- Проверим теперь, что это решение единственное, то есть если x_l – решение системы, то $x_l = p_l$. Используя индукцию и первые уравнения системы легко показать, что

$$\forall m \in \mathbb{N} \quad x_j = \sum_{k=1}^n x_k p_{k,j}(m).$$

Действительно, при $m = 1$ имеем $p_{l,j}(1) = p_{l,j}$. Предположим, что утверждение справедливо при $m = l$, подставим вместо x_k выражение из первого уравнения системы имеем

$$x_j = \sum_{k=1}^n \sum_{i=1}^n x_i p_{i,k} p_{k,j}(l) = \sum_{i=1}^n x_i p_{i,j}(l+1).$$

Осталось перейти к пределу при $m \rightarrow +\infty$:

$$x_j = \sum_{k=1}^n x_k p_j = p_j \cdot \sum_{k=1}^n x_k = p_j \cdot 1 = p_j.$$

■

Теорема 24.3. Если все элементы матрицы P строго положительны, то соответствующая цепь маркова является эргодической.

Доказательство. Обозначим

$$M_j(t) = \max_{1 \leq i \leq n} p_{i,j}(t), \quad m_j(t) = \min_{1 \leq i \leq n} p_{i,j}(t).$$

Выполнено:

$$p_{i,j}(t+1) = \sum_{k=1}^n p_{i,k} p_{k,j}(t), \quad m_j(t) \leq p_{k,j}(t) \leq M_j(t).$$

Следовательно, для любого i имеем:

$$m_j(t) = m_j(t) \sum_{k=1}^n p_{i,k} \leq p_{i,j}(t+1) \leq M_j(t) \sum_{k=1}^n p_{i,k} = M_j(t).$$

Обозначим за k, l индексы, при которых для фиксированного j достигается $M_j(t+1)$ и $m_j(t+1)$ соответственно:

$$p_{k,j}(t+1) = M_j(t+1), \quad p_{l,j}(t+1) = m_j(t+1).$$

Пользуясь предыдущим неравенством, получаем:

$$m_j(t) \leq m_j(t+1) \leq M_j(t+1) \leq M_j(t).$$

Поскольку последовательности $m_j(t)$ и $M_j(t)$ ограничены и монотонны, существуют пределы:

$$\lim_{t \rightarrow +\infty} M_j(t) = M_j, \quad \lim_{t \rightarrow +\infty} m_j(t) = m_j.$$

Для доказательства теоремы осталось показать, что

$$\lim_{t \rightarrow +\infty} M_j(t) - m_j(t) = 0.$$

Действительно, тогда $p_{i,j}(t)$ окажется зажатой между двумя последовательностями, сходящимися к одному числу, то есть будет иметь предел. Имеем

$$\begin{aligned} M_j(t+1) - m_j(t+1) &= p_{k,j}(t+1) - p_{l,j}(t+1) = \sum_{m=1}^n p_{k,m} p_{m,j}(t) - \sum_{m=1}^n p_{l,m} p_{m,j}(t) \\ &= \sum_{m=1}^n (p_{k,m} - p_{l,m}) p_{m,j}(t) \\ &= \sum_{m=1}^{n^+} (p_{k,m} - p_{l,m}) p_{m,j}(t) + \sum_{m=1}^{n^-} (p_{k,m} - p_{l,m}) p_{m,j}(t). \end{aligned}$$

\sum^+ – сумма положительных слагаемых, \sum^- – сумма отрицательных. Верно неравенство

$$m_j(t) \leq p_{m,j}(t) \leq M_j(t).$$

Поэтому:

$$M_j(t+1) - m_j(t+1) \leq M_j(t) \sum_{m=1}^{n^+} (p_{k,m} - p_{l,m}) + m_j(t) \sum_{m=1}^{n^-} (p_{k,m} - p_{l,m}).$$

Поскольку

$$0 = 1 - 1 = \sum_{m=1}^n (p_{k,m} - p_{l,m}) = \sum_{m=1}^{n^+} (p_{k,m} - p_{l,m}) + \sum_{m=1}^{n^-} (p_{k,m} - p_{l,m}), \quad \min p_{i,j} > 0$$

имеем

$$d_{l,k} = \sum_{m=1}^{n^+} (p_{k,m} - p_{l,m}) = \sum_{m=1}^{n^-} (p_{k,m} - p_{l,m}) = d < 1.$$

Тогда получаем

$$M_j(t+1) - m_j(t+1) = d \cdot (M_j(t) - m_j(t)).$$

Отсюда имеем

$$M_j(t) - m_j(t) \leq d^t \xrightarrow{n \rightarrow +\infty} 0.$$

■

Следствие 24.4. Пусть при некотором N все элементы матрицы P^N строго положительны. Тогда соответствующая цепь Маркова является эргодической.

Доказательство. Рассмотрим марковскую цепь с $Q = P^N$. Тогда по предыдущей теореме верно

$$\exists \lim_{t \rightarrow +\infty} q_{i,j}(t) = \lim_{t \rightarrow +\infty} p_{i,j}(Nt) = p_j.$$

Вычислим требуемый предел:

$$\lim_{t \rightarrow +\infty} p_{i,j}(t) = \lim_{t \rightarrow +\infty} p_{i,j}(k + Nt) = \lim_{t \rightarrow +\infty} \left(\sum_{l=1}^n p_{i,l}(k) p_{l,j}(Nt) \right) = p_j \cdot \sum_{l=1}^n p_{i,l}(k) = p_j.$$

■

24.3 Схемы блужданий

24.3.1 Блуждание по отрезку с поглощением на концах

Пусть частица движется по целым точкам отрезка $[0, N]$ по одной точке за раз. Состояние $X(t)$ определяется текущей координатой частицы. Рассмотрим следующее правило изменения состояний: если текущая точка – крайняя точка отрезка, то с вероятностью 1 частица в ней и останется. Иначе с вероятностью p частица движется на 1 ячейку вправо, и с вероятностью $q = 1 - p$ – влево. Этот процесс – марковский, со следующей матрицей переходов:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & q & 0 & p & 0 \\ 0 & \dots & 0 & q & 0 & p \\ 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Найдем $p_{x,0}(t)$ и $p_{x,N}(t)$ – вероятности поглощения за время t при условии, что $X(0) = x$. Имеем

$$\begin{aligned} p_{x,0}(t) &= P(X(t+1) = 0 \mid X(0) = x) = P(X(t+1) = 0 \mid X(1) = x+1)P(X(1) = x+1 \mid X(0) = x) \\ &\quad + P(X(t+1) = 0 \mid X(1) = x-1)P(X(1) = x-1 \mid X(0) = x) \\ &= pp_{x+1,0}(t) + qp_{x-1,0}(t). \end{aligned}$$

$p_{y,0}$ ограничена и возрастает при $t \rightarrow +\infty$, поэтому существует предел

$$\lim_{t \rightarrow +\infty} p_{y,0}(t) = p_{y,0}(\infty).$$

Совершая такой же предельный переход в предыдущем равенстве, получаем

$$p_{x,0}(\infty) = pp_{x+1,0}(\infty) + qp_{x-1,0}(\infty).$$

Кроме того,

$$p_{0,0} = p_{0,0}(\infty) = 1, \quad p_{N,0} = p_{N,0}(\infty) = 0.$$

Полученное уравнение является уравнением в конечных разностях второго порядка. При $q \neq p$ его решение имеет вид

$$p_{x,0}(\infty) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}, \quad p_{x,N}(\infty) = \frac{1 - \left(\frac{q}{p}\right)^x}{1 - \left(\frac{q}{p}\right)^N}.$$

При $q = p = \frac{1}{2}$ решение имеет вид

$$p_{x,0}(\infty) = 1 - \frac{x}{N}, \quad p_{x,N}(\infty) = \frac{x}{N}.$$

24.3.2 Блуждание по отрезку с отражением на концах

Изменим поведение частицы на концах. Пусть она не поглощается, а отражается с определенной вероятностью. Тогда матрица переходов этого процесса устроена следующим образом:

$$P = \begin{pmatrix} q & p & 0 & 0 & \dots & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & q & 0 & p & 0 \\ 0 & \dots & 0 & q & 0 & p \\ 0 & \dots & 0 & 0 & q & p \end{pmatrix}.$$

Утверждение 24.5. При $p \in (0, 1)$ $\exists N$: P^N – эргодическая матрица.

Доказательство. Пусть $i < j$, $i + j < N$. Рассмотрим траекторию ω :

$$\underbrace{i \rightarrow (i-1) \rightarrow \dots \rightarrow 1}_{i \text{ шагов}} \rightarrow \underbrace{0 \rightarrow \dots \rightarrow 0}_{N-(i+j) \text{ шагов}} \rightarrow \underbrace{1 \rightarrow 2 \rightarrow \dots \rightarrow j}_{j \text{ шагов}}.$$

Вычислим вероятность такой траектории:

$$p_{i,j}(\omega) = p^i q^{N-i-j} p^j > 0.$$

Кроме того, очевидно, что $p_{i,j}(N) > p_{i,j}(\omega) > 0$. Аналогично можно показать, что $p_{i,j}(N) > 0$ и в случае $i \geq j$. ■

Получается, что построенная цепь является эргодической. Раз так, существуют пределы

$$p_j = \lim_{t \rightarrow +\infty} p_{i,j}(t),$$

которые можно определить из системы уравнений

$$p_j = \sum_{i=0}^N p_i p_{i,j}, \quad \sum_{j=0}^N p_j = 1.$$

Решая эту систему методом индукции в нашем случае, получаем

$$p_1 = \frac{p}{q} p_0, \quad p_2 = \frac{p^2}{q^2} p_0, \quad \dots$$

p_0 определим из условия нормировки:

$$1 = \sum_{i=0}^N p_i = p_0 \sum_{i=0}^N \frac{p^i}{q^i} \Rightarrow p_0 = \frac{1}{\sum_{i=0}^N \frac{p^i}{q^i}} = \frac{\frac{p}{q} - 1}{\left(\frac{p}{q}\right)^{N+1} - 1}.$$

Устремляя $N \rightarrow +\infty$, получим результат для счётных марковских цепей.

$$p_k = \frac{p^k}{q^k} \left(1 - \frac{p}{q}\right).$$