Математическая статистика

15 июня 2020 г.

Содержание

1 Постановка задач математической статистики **3**

1.1 Задачи теории вероятностей 3

1.2 Задачи математической статистики 3

2 Частота как оценка вероятности события и её свойства. Построение доверительного интервала для вероятности события на базе асимптотической нормальности частоты. **5**

3 Постановка выборочной статистической модели. Точечная оценка параметра и характеристики. **7**

4 Функции потерь и функции риска, состоятельность оценки характеристики, достаточное условие для состоятельности оценки. **8**

5 Вид квадратичного риска в случае одномерной характеристики. **10**

6 Постановка задачи доверительного оценивания, доверительный интервал. **11**

7 Определение несмещенности и асимптотической нормальности оценки характеристики. Построение доверительного интервала для характеристики на базе асимптотической нормальности ее оценки. **12**

8 Постановка задачи проверки гипотез **13**

9 Ошибки первого и второго рода и их вероятности как критерий качества критерия (теста) проверки гипотез. Подход Неймана-Пирсона. **14**

10 Эмпирическая функция распределения (ЭФР). Построение, свойства ЭФР при фиксированном значении аргумента (использовать свойства частоты). **15**

11 Свойства ЭФР в целом. Расстояние Колмогорова, Смирнова. Теоремы Гливенко-Кантелли, Колмогорова, Мизеса-Смирнова. Построение доверительной полосы для функции распределения. **16**

12 Критерии согласия Колмогорова и Мизеса-Смирнова. 18
12.1 Критерий согласия Колмогорова 18
12.2 Критерий Мизеса-Смирнова 18
12.3 Прикладной алгоритм 19

13 Выборочный метод построения оценок одномерных характеристик. Асимптотическая нормальность оценки. Построение асимптотического доверительного интервала на базе асимптотической нормальности. 20
13.1 Описание выборочного метода 20
13.2 Асимптотическая нормальность, свойства асимптотической нормальности оценок 20

14 Основные выборочные оценки и их свойства. Выборочное математическое ожидание. Выборочная дисперсия. Выборочные моменты. Выборочные медиана и квантили. Выборочные оценки ковариации и коэффициента корреляции. 22
14.1 Выборочное среднее / М.О. 22
14.2 Выборочная дисперсия 23
14.3 Несмещенная выборочная дисперсия 23
14.4 Выборочные моменты 24
14.4.1 Выборочные начальные моменты 24
14.4.2 Выборочные центральные моменты 24
14.5 Выборочная медиана 24
14.6 Выборочная ковариация и корреляция 25
14.6.1 Выборочная ковариация 25
14.7 Выборочная корреляция 26

15 Метод моментов и его свойства. 27
15.1 Идея метода подстановки 27
15.2 Метод моментов 27

16 Метод максимального правдоподобия и его свойства. 29

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 $X : \Omega \to \mathbb{R}^n$.

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

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

Для произвольного A.

Определение. Пусть

$$\overline{X} = \langle X_1, \ldots, X_n \rangle.$$

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 $\hat{g} : \mathcal{X} \to \Theta$. 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 проеркой гипотезы назовем построение *теста* (*критерия*), то есть отображения
  $$\varphi : \mathcal{X} \to N.$$
  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

$$\overline{\xi}_n = \frac{1}{n}\sum_{i=1}^{n} \xi_n = \frac{k_n}{n}.$$

Тогда

$$\overline{\xi}_n \underset{n\to+\infty}{\rightrightarrows} p.$$

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

$$Y_n = \frac{\overline{X}_n - E\left(\overline{X}_n\right)}{\sigma\left(\overline{X}_n\right)}.$$

Верно:

$$F_{Y_n} \underset{\mathbb{R}}{\rightrightarrows} F_{N(0,1)}.$$

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:

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

Это буквально означает:

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

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

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

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

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

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

Определение. $\hat{g}_n$ называется *состоятельной*, если $\hat{g}_n \underset{P}{\Longrightarrow} g(\theta)$ при $n \to \infty$.

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

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

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

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

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

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

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

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 являтся случайной величиной.

Определение. *Риском* называется функция

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

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

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

Соответствующий риск будет вероятностью отклонения:

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

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

$$\forall \delta > 0 \; R^\delta(\hat{g}_n, \theta) = P_{n,\theta}(\|\hat{g}_n - g(\theta)\| \geqslant \delta) \xrightarrow[n \to +\infty]{} 0.$$

Или, что то же самое:

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

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)
$R_2(\hat{g}_n, \theta) \xrightarrow[n\to+\infty]{} 0 \Longrightarrow$ оценка состоятельна.

Доказательство.

$$\begin{aligned} \forall \delta > 0 \ R^\delta(\hat{g}_n, \theta) = P(\|\hat{g}_n - g(\theta)\| \geqslant \delta) &= P(\|\hat{g}_n - g(\theta)\|^2 \geqslant \delta^2) \\ &\leqslant \frac{E_\theta(\|\hat{g}_n - g(\theta)\|^2)}{\delta^2} = \frac{R_2(\hat{g}_n, \theta)}{\delta^2} \xrightarrow[n\to+\infty]{} 0. \end{aligned}$$

■

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 $\widetilde{G}_\gamma \subseteq G = g(\Theta)$, обладающая свойством:

$$\forall \theta \in \Theta \; P_\theta(g(\theta) \in \widetilde{G}_\gamma) \geqslant \gamma.$$

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* $\widetilde{G}_{n,\gamma}$ задается условием:

$$\forall \theta \in \Theta \; \lim P_{n,\theta}(g(\theta) \in \widetilde{G}_{n,\gamma}) \geqslant \gamma.$$

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:

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

Это буквально означает:

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

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

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

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

8 Постановка задачи проверки гипотез

Определение. *Гипотезой* называется множесто 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:

$$\varphi(\mathcal{X}) : \mathcal{X}_n \to \{0, 1\}.$$

При этом:

- $\mathcal{X}_{n,0}$ называют *допустимым множеством*.
- $\mathcal{X}_{n,1}$ называют *критическим множеством*.
- $\mathcal{X}_{n,0} \sqcup \mathcal{X}_{n,1} = \mathcal{X}_n$.

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

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

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*:

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

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

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

Определение. β называют *вероятностью ошибки II рода*:

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

Определение. *Мощностью теста* называют следующую величину:

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

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:

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

Среди этих тестов выбирается тест с минимальным значением β.
В асимптотических задачах ограничения накладываются на предельные значения.

10 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 фкисированном $t = t_0$:

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

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

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

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

11 Свойства ЭФР в целом. Расстояние Колмогорова, Смирнова. Теоремы Гливенко-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*:

$$\rho_2^2(F_n, F_x) = \int\limits_{\mathbb{R}} (F_n(t) - F_x(t))^2 \, \mathrm{d}F_x(t).$$

Теорема 11.1. (Гливенко-Кантелли)
Пусть $\mathcal{F}$ – множество функций распределения. Тогда $\forall F_x(t) \in \mathcal{F}$ с вероятностью 1 справедливо предельное неравенство:

$$\rho_\infty(F_n, F_x) \xrightarrow[n\to\infty]{} 0.$$

Так как $\rho_2 \leqslant \rho_\infty$, 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 11.2. (Колмогоров)

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

Теорема 11.3. (Мизес, Смирнов)

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

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

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

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:

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

Утверждение 11.4.
$$P_x(F_n^-(t) \leqslant F_x(t) \leqslant F_n^+(t)) \xrightarrow[n\to\infty]{} \gamma.$$

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

$$P_x\big(F_n^-(t) \leqslant F_x(t) \leqslant F_n^+(t)\big) = P_x\left(F_n(t) - \frac{u_\gamma}{\sqrt{n}} \leqslant F_x(t) \leqslant F_n(t) + \frac{u_\gamma}{\sqrt{n}}\right) \overset{\forall t}{=}$$
$$\overset{\forall t}{=} P_x\big(\sqrt{n}|F_x(t) - F_n(t)| \leqslant u_\gamma\big) \overset{\forall t}{=}$$
$$\overset{\forall t}{=} P_x\left(\sqrt{n}\sup_t |F_x(t) - F_n(t)| \leqslant u_\gamma\right) \xrightarrow{\text{th. Колмогорова}} \mathcal{K}(u_\gamma) = \gamma.$$

■

12 Критерии согласия Колмогорова и Мизеса-Смирнова.

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

Поставим задачу проверки согласия:

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

12.1 Критерий согласия Колмогорова

Определим тестовую статистику:

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

По th. Колмогорова:

$$P(L(\overline{X}) < z) \xrightarrow[n\to\infty]{} \mathcal{K}(z),$$

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:

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

12.2 Критерий Мизеса-Смирнова

Определим тестовую статистику:

$$L(\overline{X}) = \sqrt{n}\rho_2^2(F_0, F_n).$$

По th. Мизеса-Смирнова:

$$P(L(\overline{X}) < z) \xrightarrow[n\to\infty]{} \mathcal{S}(z),$$

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:

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

12.3 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 – оклонить.

13 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

13.1 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:

Y_i	$x_{(1)}$	$x_{(2)}$	...	$x_{(n)}$
p_i	$1/n$	$1/n$	...	$1/n$

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

$$x_{(1)} \leqslant x_{(2)} \leqslant ... \leqslant x_{(n)}.$$

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.

13.2 Асимптотическая нормальность, свойства асимптотической нормальности оценок

Определение (для одномерного параметра θ, $g(\theta)$) Последовательность оценок $\hat{g}_n$ характеристики $g(\theta)$ def асимптотически нормальной с ас. дисперсией $\Delta^2(\theta) > 0$, если сл. в. $Y_n = \sqrt{n}(\hat{g}_n - g(\theta))$ сходится по $P_{n,x}$ - распределению к нормальной сл. в. Y с нулевым средним и дисперсией $\Delta^2(\theta)$

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

Перепишем (1)

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

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

$\Delta(\theta)$ - def нормирующим множителем и $P_{n,\theta}(\frac{Y_n}{\Delta(\theta)} < t) \xrightarrow[n\to\infty]{} F_{N(0,1)}(t) \Rightarrow P_{n,x}(|\hat{g}_n - g(\theta)| < \frac{T(\Delta(\theta))}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 2\Phi(T) - 1 = \gamma \Rightarrow T = \frac{1+\gamma}{2}$ - 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

14 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.

14.1 Выборочное среднее / М.О.

Определение. Случайную величину $\overline{X}_n = EY = \sum\limits_{i=1}^{n} X_{(i)} \cdot \frac{1}{n} = \frac{1}{n}\sum\limits_{j=1}^{n} X_j$ 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 законму больших чисел:

$$\overline{X}_n \xrightarrow[n\to\infty]{p} EX, \text{ если } \exists EX.$$

  Поэтому выборочное среднее является *состоятельной* оценкой EX.

- Выборочное среднее является *несмещенной* оценкой EX:

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

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

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

  или

$$F_{Y_n}(t) \underset{n\to\infty}{\overset{\mathbb{R}}{\rightrightarrows}} F_Y(t).$$

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

14.2 Выборочная дисперсия

Определение. Случайную величину $S_n^2 = D(X_{[n]}) = \sum\limits_{i=1}^{n}(X_{(i)} - EY)^2 \cdot \frac{1}{n} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$ 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:

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

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

$$\begin{aligned} &E(X^2) = DX,\ E_x(\overline{X}_n) = EX = 0 \\ &\Rightarrow E_X(\overline{X}_n^2) = D_X(\overline{X}_n) + (E_x(\overline{X}_n))^2 = \frac{DX}{n} \\ &\Rightarrow E_X S_n^2 = E_X(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_i^2) = \frac{1}{n}nE(X^2) - \frac{DX}{n} = DX - \frac{DX}{n} = \frac{n-1}{n}DX. \end{aligned}$$

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

14.3 Несмещенная выборочная дисперсия

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

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

Несмещенная выборочная дисперсия является выборочной точечной оценкой DX.

Свойства.

- *Состоятельность* следует из состоятельности S_n^2:

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

- *Несмещенность* очевидна из доказательства смещенности выборочного среднего.

- Несмещенная оценка дисперсии является *асимптотически нормальной* оценкой DX – *без доказательства*.

14.4 Выборочные моменты

14.4.1 Выборочные начальные моменты

Определение. *Выборочным начальным моментом порядка* k называется статистика:

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

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

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

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 $\Delta^2 = E_x(Z^2) - (E_x(Z))^2$.

14.4.2 Выборочные центральные моменты

Определение. *Выборочным центральным моментом порядка* k называется статистика:

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

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

14.5 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:

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

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

Определение. Пусть $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ – упорядоченная выборка (вариационный ряд), тогда *выборочной vедианой* 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

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

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

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

- Следовательно, выборочная медиана является *состоятельной* оценкой $med(X)$.
- Также видно, что выборочная медиана является *асимптотически нормальной* оценкой $med(x)$ с асимптотической дисперсией $\Delta^2 = \frac{1}{r f_X^2(t_0)}$.
- Выборочная медиана является $\sqrt{n}$*-несмещенной* оценкой $med(X)$. То есть:

$$\sqrt{n} b_{n,\theta}(med_n) = \sqrt{n}(E_x(med_n) - med(X)) \xrightarrow[n\to\infty]{} 0.$$

14.6 Выборочная ковариация и корреляция

Выборочная ковариация и корелляция 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.

14.6.1 Выборочная ковариация

Определение. *Выборочной ковариацией* называется статистика:

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

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*:

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

14.7 Выборочная корреляция

Определение. *Выборочной корреляцией* X и Y называется статистика:

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

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

Свойства.

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

15 Метод моментов и его свойства.

15.1 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.

15.2 Метод моментов

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

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

Здесь $H\colon \mathbb{R} \to \mathbb{R}$ – известная нам функция. Метод состоит в том, чтобы *заменить* математическое ожидание его выборочной оценкой, то есть в качестве оценки параметра $\hat{\theta}\big(X^{(n)}\big)$ 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 характерисика $g : \mathbb{R}^m \to \mathbb{R}^k$ удовлетворяет следующим свойствам:

- Система уравнений относительно θ:

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

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

- Система уравнениий 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:

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

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

Теорема 15.1. (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

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

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:

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

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

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

Где

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

■

16 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:

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

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 нызвается $\alpha(n)$-несмещенной, если

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

Теорема 16.1. (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 $k \leqslant 3$.
- Выполнены неравенства

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

  Причем G_k суммируемы и

  $$\sup_{\theta\in\Theta} \int\limits_{\mathbb{R}} G_3(x) f(x, \theta)\,\mathrm{d}x < +\infty.$$

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

Тогда соответстующая оценка максимального правдоподобия обладает свойствами:

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