Статистика

Конспект практических занятий

Мат-Мех, ПМИ, СМ–СМ

$Revision: ... (NEG: UNDER RECONSTRUCTION) $

СтатМод

Оглавление

Часть I.

Оценки характеристик и параметров распределения

1. Выборка и эмпирическая случайная величина

Пусть $\xi \sim \mathcal{P}$ — случайная величина с распределением $\mathcal{P}$.

Определение. *Повторной независимой выборкой* объема n *(до эксперимента)* называется набор

$$\mathbf{x} = (x_1, \ldots, x_n), \quad x_i \sim \mathcal{P}\ \forall i \in 1:n,\ x_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp x_n$$

независимых в совокупности одинаково распределенных случайных величин с распределением $\mathcal{P}$.

Определение. *Повторной независимой выборкой* объема n *(после эксперимента)* называется набор реализаций, т.е. конкретных значений ξ, случайных величин x_i:

$$\mathbf{x} = (x_1, \ldots, x_n), \quad x_i \in \operatorname{supp} \xi\ \forall i \in 1:n.$$

Определение. *Эмпирической случайной величиной* $\hat{\xi}_n$ называется случайная величина с дискретным распределением

$$\hat{\xi}_n \sim \hat{\mathcal{P}}_n : \begin{pmatrix} x_1 & \ldots & x_n \\ 1/n & \ldots & 1/n \end{pmatrix}.$$

Замечание. Подходящее определение выбирается по контексту.

Если ξ имеет дискретное распределение, то выборку можно *сгруппировать*; тогда получим случайную величину $\hat{\xi}_m$ с распределением

$$\hat{\mathcal{P}}_m : \begin{pmatrix} x_1^* & \ldots & x_m^* \\ \omega_1 & \ldots & \omega_m \end{pmatrix} \quad \omega_i = \frac{\nu_i}{n},$$

где x_i^* — уникальные значения из выборки $\mathbf{x}$, а ν_i — число x_i^* в $\mathbf{x}$ (т.н. «абсолютная частота»; тогда ω_i — «относительная частота»). В противном случае, можно разбить интервал всевозможных значений выборки на m подынтервалов: $\{[e_0, e_1), \ldots, [e_{m-1}, e_m)\}$ и считать число наблюдений $\nu_i = \nu_i[e_{i-1}, e_i)$, попавших в интервал.

Следствие. *По ЗБЧ (теореме Бернулли),*

$$\omega_i \xrightarrow{\mathsf{P}} p_i = \mathsf{P}(e_{i-1} \leq \xi < e_i),$$

т.е. относительная частота является хорошей оценкой вероятности на больших объемах выборки.

2. Виды признаков

Виды признаков случайной величины $\xi : \big(\Omega, \mathcal{F}, \mathsf{P}\big) \to \big(V, \mathfrak{A}\big)$ характеризуются тем, что из себя представляет множество V и что можно делать с его элементами.

Количественные признаки: $V \subset \mathbb{R}$

По типу операций:

- Аддитивные: заданы, т.е. имеют смысл в контексте данного признака, операции $+, -$
- Мультипликативные: заданы операции $\cdot, /$; признак принимает не отрицательные значения.

По типу данных:

- Непрерывные
- Дискретные

Порядковые признаки V — упорядоченное множество, определены отношения $>, =$.

Качественные признаки на V заданы отношения $=, \neq$

Пример. Цвет глаз, имена, пол.

3. Характеристики распределений и метод подстановки

Определение. *Статистика* — измеримая функция от выборки.

Обобщением статистики является понятие характеристики.

Определение. *Характеристика* — функционал от распределения:

$$T : \{\mathcal{P}\} \to V.$$

Где V — измеримое пространство, чтобы на нём можно было завести σ-алгебру.

Замечание. Чаще всего, $V = \mathbb{R}$.

Определение. Выделяют *генеральные* характеристики $T(\mathcal{P}) =: \theta$ и *выборочные* характеристики $T(\hat{\mathcal{P}}_n)$.

Определение. *Оценка* — выборочная характеристика $T(\hat{\mathcal{P}}_n) =: \hat{\theta}_n$, не зависящая от генеральной характеристики θ.

Следствие. *Выражения для вычисления генеральных и выборочных характеристик отличаются только используемыми мерами ($\mathcal{P}$ и $\hat{\mathcal{P}}_n$ соответственно).*

Определение. Пусть $\hat{\mathcal{P}}_n$ — распределение эмпирической случайной величины. Тогда *эмпирическая функция распределения* есть

$$\widehat{\mathrm{cdf}}_\xi(x) = \mathrm{cdf}_{\hat{\xi}_n}(x) = \hat{\mathcal{P}}_n((-\infty, x)) = \int_{-\infty}^{x} \mathrm{d}\hat{\mathcal{P}}_n = \sum_{x_i : x_i \le x} \frac{1}{n} = \frac{|\{x_i \in \mathbf{x} : x_i < x\}|}{n}.$$

Утверждение. Пусть $\widehat{\mathrm{cdf}}_\xi$ — эмпирическая функция распределения, cdf_ξ — функция распределения ξ. Тогда, по теореме Гливенко-Кантелли,

$$\sup_x \left| \widehat{\mathrm{cdf}}_\xi(x) - \mathrm{cdf}_\xi(x) \right| \xrightarrow{\text{a.s.}} 0.$$

Более того, если cdf_ξ непрерывна, эта сходимость имеет порядок $1/\sqrt{n}$ по теореме Колмогорова:

$$\sqrt{n} \sup_{x \in \mathbb{R}} \left| \widehat{\mathrm{cdf}}_\xi(x) - \mathrm{cdf}_\xi(x) \right| \xrightarrow{\mathrm{d}} \mathcal{P}_{\text{K.S.}},$$

где $\mathcal{P}_{\text{K.S.}}$ — распределение Колмогорова-Смирнова.

Замечание. Поскольку $\widehat{\mathrm{cdf}}_\xi(x) = \omega_x$, где ω_x — частота попадания наблюдений в интервал в $(-\infty, x)$, а $\mathrm{cdf}_\xi(x) = \mathsf{P}\left(\xi \in (-\infty, x)\right)$ — вероятность того же события, то можно применить теорему Бернулли (ЗБЧ):

$$\widehat{\mathrm{cdf}}_\xi(x) \xrightarrow{\mathsf{P}} \mathrm{cdf}_\xi(x).$$

Следствие. *Значит, при достаточно больших n, в качестве интересующей характеристики θ распределения $\mathcal{P}$ можем брать ее оценку $\hat{\theta}_n$ — аналогичную характеристику $\hat{\mathcal{P}}_n$.*

4. Характеристики распределений и их оценки

Определение. Генеральные и соответствующие им выборочные характеристики *k-го момента* и *k-го центрального момента*:

$$\mathsf{m}_k = \int_{\mathbb{R}} x^k \, \mathrm{d}\mathcal{P} \qquad \hat{\mathsf{m}}_k = \int_{\mathbb{R}} x^k \, \mathrm{d}\hat{\mathcal{P}}_n = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$$\mathsf{m}_k^{(0)} = \int_{\mathbb{R}} (x - \mathsf{m}_1)^k \, \mathrm{d}\mathcal{P} \qquad \hat{\mathsf{m}}_k^{(0)} = \int_{\mathbb{R}} (x - \hat{\mathsf{m}}_1)^k \, \mathrm{d}\hat{\mathcal{P}}_n = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mathsf{m}}_1)^k.$$

4.1. Характеристики положения

В качестве характеристики положения выделяется 1-й момент — математическое ожидание и *выборочное среднее*:

$$\mathsf{m}_1 = \mathsf{E}\xi, \qquad \hat{\mathsf{m}}_1 =: \bar{x} = \widehat{\mathsf{E}\xi} = \mathsf{E}\hat{\xi}_n.$$

Замечание. В случае мультипликативных признаков можно посчитать среднее геометрическое; часто логарифмируют и считают среднее арифметическое.

Определение. Пусть $p \in [0, 1]$ и $\mathrm{cdf} = \mathrm{cdf}_{\mathcal{P}}$. *$p$-квантилью* (квантилью уровня p) называется

$$\mathrm{qnt}_{\mathcal{P}}(p) =: z_p = \sup\left\{z : \mathrm{cdf}(z) \le p\right\}.$$

Квартиль есть квантиль уровня, кратного $1/4$; дециль — $1/10$; перцентиль — $1/100$.

Замечание. sup берется для учета случая не непрерывных функций распределения.

Определение. *Медиана* есть $1/2$-квантиль:

$$\mathrm{med}\,\xi = z_{1/2}.$$

Определение. Мода ($\mathrm{mode}\,\xi$) есть точка локального максимума плотности.

По методу подстановки можем получить аналогичные выборочные характеристики.

Определение. *Выборочная p-квантиль* есть такая точка $\hat{z}_p$, что она больше по значению $|\mathbf{x}| \cdot p = np$ точек из выборки:

$$\hat{z}_p = \sup\left\{z : \widehat{\mathrm{cdf}}_\xi(z) \le p\right\} = x_{(\lfloor np \rfloor + 1)}.$$

Определение. *Выборочная медиана* упорядоченной выборки $\mathbf{x} = \left(x_{(1)}, \ldots, x_{(n)}\right)$ есть

$$\hat{z}_{1/2} = \widehat{\mathrm{med}} = \begin{cases} x_{(k+1)} & n = 2k+1 \\ \dfrac{x_{(k)} + x_{(k+1)}}{2} & n = 2k \end{cases}$$

Определение. *Выборочная мода* ($\widehat{\mathrm{mode}}$) есть значение из выборки, которое чаще всего встречается.

4.2. Характеристики разброса

В качестве характеристики разброса выделяется 2-й центральный момент — дисперсия и выборочная дисперсия:

$$\mathsf{m}_2^{(0)} = \mathsf{D}\xi \qquad \hat{\mathsf{m}}_2^{(0)} =: s^2 = \widehat{\mathsf{D}\xi} = \mathsf{D}\hat{\xi}_n = \begin{cases} \mathsf{E}\left(\hat{\xi}_n - \mathsf{E}\hat{\xi}_n\right)^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ \mathsf{E}\hat{\xi}_n^2 - \left(\mathsf{E}\hat{\xi}_n\right)^2 = \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2. \end{cases}$$

Замечание. Если среднее $\mathsf{E}\xi = \mu$ известно, то дополнительно вводится

$$s_\mu^2 := \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 \\ \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^2\right) - \mu^2. \end{cases}$$

Пример (Оценка дисперсии оценки мат. ожидания)**.** Пусть строится оценка мат. ожидания $\bar{x}$. Может интересовать точность построенной оценки. Вычислим дисперсию теоретически, после чего оценим точность по выборке:

$$\mathsf{D}\bar{x} = \mathsf{D}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{D}x_i = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{D}\xi = \frac{\mathsf{D}\xi}{n},$$

откуда

$$\widehat{\mathsf{D}\bar{x}} = \frac{s^2}{n}.$$

Пример (Дисперсия оценки дисперсии)**.** См. по ссылке[1].

Определение (Энтропия)**.** Количество информации, необходимое для выявления объекта из n-элементного множества вычисляется по *формуле Хартли*:

$$H = \log_2 n$$

(множество это следует итеративно разбивать пополам, откуда и оценка). Пусть теперь множество не равновероятно, т.е. задано дискретное распределение

$$\mathcal{P}_\xi : \begin{pmatrix} x_1 & \dots & x_n \\ p_1 & \dots & p_n \end{pmatrix}.$$

Тогда количество информации $H(\xi)$, которую нужно получить, чтобы узнать, какой исход эксперимента осуществлен, вычисляется по формуле Шеннона и называется *энтропией*:

$$H(\xi) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}.$$

Замечание. В случае равномерного дискретного распределения, конечно, $H = H(\xi)$.

Определение. *Выборочное стандартное отклонение* есть

$$\mathrm{SD} := \sqrt{\widehat{\mathsf{D}\xi}} = s.$$

Это показатель разброса случайной величины; показатель того, насколько элементы выборки отличаются от выборочного среднего по значению.

[1] http://mathworld.wolfram.com/SampleVarianceDistribution.html

SD позволяет оценивать стандартное отклонение распределения ξ.

Пусть $\hat{\theta}_n$ — статистика. Она имеет какое-то своё распределение, стандартное отклонение которого можно также оценить.

Определение. *Стандартная ошибка* оценки есть

$$\mathrm{SE}(\hat{\theta}) := \sqrt{\widehat{\mathsf{D}\hat{\theta}}}.$$

Это показатель разброса оценки случайной величины.

Замечание. В частном случае $\theta = \mathsf{E}\xi$, $\hat{\theta} = \bar{x}$ получаем *выборочную стандартную ошибку среднего*

$$\mathrm{SE} := \mathrm{SE}(\bar{x}) = \sqrt{\widehat{\mathsf{D}\bar{x}}} = \sqrt{\frac{\widehat{\mathsf{D}\xi}}{n}} = \frac{s}{\sqrt{n}}.$$

Это, в свою очередь, показатель того, насколько выборочное среднее отличается от истинного.

Пусть $c_\gamma = \mathrm{qnt}_{\mathrm{N}(0,1)}\,\gamma$.

Пример (С мостом и машинами)**.** При возведении моста требуется, чтобы под ним могли проехать, условно, 95% машин. Чтобы эту высоту вычислить, достаточно собрать выборку высоты кузова проезжающих машин. Тогда нахождение искомой величины можно наглядно представить как выбор такой квантили гистограммы выборки, что суммирование соответствующих вероятностей даст 0.95. В предположении, что выборка из нормального распределения, с более устойчивой оценкой квантили, интервал будет иметь вид

$$\left(\bar{\mathbf{x}} \pm \mathrm{SD} \cdot c_\gamma\right).$$

SE как показатель разброса среднего использовать по смыслу нельзя.

Пример (С паромом)**.** Число машин, которое способен перевезти паром, есть Грузоподъемность/$\mathsf{E}\xi$, где ξ — вес машины. Поскольку оценка $\bar{x}$ всегда считается с погрешностью относительно истинного значения, интервал допустимого числа машин будет иметь вид

$$\frac{\text{Грузоподъемность}}{\bar{x} \pm \mathrm{SE} \cdot c_\gamma}.$$

4.3. Анализ характера разброса

Определение. *Коэффициент асимметрии Пирсона* («скошенности»[2])

$$\gamma_3 = \mathsf{A}\xi = \frac{\mathsf{m}_3^{(0)}}{\sigma^3}.$$

Замечание. Не зависит от линейных преобразований.

Замечание. Старое определение скошенности было $\frac{\mathsf{E}\xi - \mathrm{med}\,\xi}{\sigma}$.

Замечание. Типичный случай соответствует тому, что при положительном коэффициенте асимметрии 'хвост вправо'.

Определение. *Коэффициент эксцесса* («крутизны», «kurtosis»):

$$\gamma_4 = \mathsf{K}\xi = \frac{\mathsf{m}_4^{(0)}}{\sigma^4} - 3.$$

Замечание. Величина $\mathsf{m}_4^{(0)}/\sigma^4 = 3$ соответствует стандартному нормальному распределению. Так что можно сравнивать выборку и γ_4 $\mathrm{N}(0,1)$.

Замечание. Положительный коэффициент эксцесса соответствует медленному убыванию на концах отрезка. Причём, так как распределение стандартизуется, имеется в виду убывание на хвостах, которое медленнее по порядку (!), чем убывание на хвостах у нормального распределения. Например, сравните e^{-x^2}, $e^{-x^2/10}$ и e^{-10*x}. Часто говорят об островершинности при положительном эксцессе, но это просто вторая сторона скорости убывания на хвостах.

[2] «Skewness».

4.4. Характеристики зависимости

Определение. Пусть $(\xi_1, \xi_2) \sim \mathcal{P}$ и $(x_1, y_1), \ldots, (x_n, y_n) \sim \mathcal{P}(du \times dv)$. Тогда можно записать две другие важные характеристики: *ковариацию* и *коэффициент корреляции*:

$$\operatorname{cov}(\xi_1, \xi_2) = \iint_{\mathbb{R}^2} (u - \mathsf{m}_1(\xi_1))(v - \mathsf{m}_1(\xi_2))\mathcal{P}(du \times dv) \qquad \widehat{\operatorname{cov}}(\mathbf{x}, \mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\operatorname{cor}(\xi_1, \xi_2) = \frac{\operatorname{cov}(\xi_1, \xi_2)}{\sigma_{\xi_1}\sigma_{\xi_2}} \qquad \widehat{\operatorname{cor}}(\mathbf{x}, \mathbf{y}) = \frac{\widehat{\operatorname{cov}}(\mathbf{x}, \mathbf{y})}{s(\mathbf{x})s(\mathbf{y})}.$$

5. Точечная оценка параметров распределения

5.1. Метод подстановки

Метод подстановки заключается в подстановке вместо неизвестного теоретического распределения известного эмпирического распределения. Например, вас интересует некоторая характеристика $f(\xi)$, а вы в качестве оценки предлагаете $\widehat{f(\xi)} = f(\xi_n)$, где ξ_n — эмпирическая случайная величина.

5.2. Метод моментов

Пусть $\mathcal{P}(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)^{\mathsf{T}}$ — параметрическая модель. Найдем оценки для параметров $\hat{\theta}_i$, $i \in \overline{1:r}$, для чего составим и решим систему уравнений:

$$\begin{cases} \mathsf{E}g_1(\xi) = \phi_1(\theta_1, \ldots, \theta_r) \\ \qquad \vdots \\ \mathsf{E}g_r(\xi) = \phi_r(\theta_1, \ldots, \theta_r) \end{cases} \implies \begin{cases} \theta_1 = f_1(\mathsf{E}g_1(\xi), \ldots, \mathsf{E}g_r(\xi)) \\ \ \vdots \\ \theta_r = f_r(\mathsf{E}g_1(\xi), \ldots, \mathsf{E}g_r(\xi)). \end{cases}$$

Примем

$$\theta_i^* = f_i(\hat{\mathsf{E}}g_1(\xi), \ldots, \hat{\mathsf{E}}g_r(\xi)).$$

Часто, $g_i(\xi) = \xi^i$. Или, еще чаще, $g_1(\xi) = \xi$ и $g_i(\xi) = (\xi - \mathsf{E}\xi)^i$, $i > 1$, так как для таких моментов обычно известны формулы.

Замечание. Случается, что решение находится вне пространства параметров. На практике, если пространство параметров компактное, можно взять точку, ближайшую к полученной оценке. Однако это свидетельствует о том, что модель плохо соответствует данным.

Пример 5.1 ($r = 1$)**.** $\xi \sim \mathrm{U}(0, \theta)$.

- Оценка по 1-му моменту: $g(\xi) = \xi$ и

$$\mathsf{E}\xi = \int_0^\theta \frac{1}{\theta} x \, \mathrm{d}x = \frac{1}{\theta} \left. \frac{x^2}{2} \right|_0^\theta = \frac{\theta}{2} \implies \theta = 2\mathsf{E}\xi, \ \theta^* = 2\bar{x}.$$

- Оценка по k-му моменту: $g(\xi) = \xi^k$ и

$$\mathsf{E}\xi^k = \frac{1}{\theta} \int_0^\theta x^k \, \mathrm{d}x = \frac{1}{\theta} \left. \frac{x^{k+1}}{k+1} \right|_0^\theta = \frac{\theta^k}{k+1} \implies \theta^* = \sqrt[k]{(k+1)\frac{1}{n}\sum_{i=1}^n x_i^k}.$$

Пример 5.2 ($r = 1$)**.** Пусть $\xi \sim \mathrm{Exp}(\lambda)$. Тогда $\mathsf{E}\xi = \lambda$ и $\bar{x} = \lambda$.

5.3. Метод оценки максимального правдоподобия

Пусть $\mathcal{P}_\xi(\boldsymbol{\theta}),\ \boldsymbol{\theta} = (\theta_1, \dots, \theta_r)^\mathsf{T}$ — параметрическая модель.

Определение. Пусть

$$\mathsf{P}(\mathbf{y} \mid \boldsymbol{\theta}) = \begin{cases} \mathsf{P}_{\boldsymbol{\theta}}(x_1 = y_1, \dots, x_n = y_n) & \mathcal{P}_\xi(\boldsymbol{\theta}) \text{ дискретно;} \\ p_{\boldsymbol{\theta}}(\mathbf{y}) & \mathcal{P}_\xi(\boldsymbol{\theta}) \text{ абсолютно непрерывно.} \end{cases}$$

Тогда *функция правдоподобия* определяется как значение распределения выборки (плотности в непрерывном случае и вероятности значений в дискретном) с подстановкой выборки вместо аргумента:

$$\mathsf{L}(\boldsymbol{\theta} \mid \mathbf{x}) = \mathsf{P}(\mathbf{x} \mid \boldsymbol{\theta}).$$

Пример 5.3. Пусть $\xi \sim \mathrm{N}(\mu, \sigma^2)$. По независимости x_i, $p_{\boldsymbol{\theta}}(\mathbf{x})$ распадается в произведение:

$$\mathsf{L}(\boldsymbol{\theta} \mid \mathbf{x}) = p_{\boldsymbol{\theta}}(\mathbf{x}) = \prod_{i=1}^n p_{\boldsymbol{\theta}}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right\}.$$

Пример 5.4. $\xi \sim \mathrm{Pois}(\lambda)$,

$$\mathsf{P}(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda} \implies \mathsf{L}(\boldsymbol{\theta} \mid \mathbf{x}) = \prod_{i=1}^n \frac{1}{x_i!}\lambda^{x_i}e^{-\lambda} = \frac{1}{\prod_{i=1}^n x_i!}\lambda^{n\bar{x}}e^{-n\lambda}.$$

Утверждение. Пусть $\mathbf{x}$ — выборка. В качестве оценки максимального правдоподобия[1] $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ следует взять

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \log \mathsf{L}(\boldsymbol{\theta} \mid \mathbf{x}).$$

Пример. $\xi \sim \mathrm{Pois}(\lambda)$.

$$\ln \mathsf{L}(\lambda \mid \mathbf{x}) = -\sum_{i=1}^n \ln(x_i!) - n\lambda + n\bar{x}\ln\lambda \implies \frac{\partial \log \mathsf{L}(\lambda \mid \mathbf{x})}{\partial \lambda} = -n + \frac{n\bar{x}}{\lambda}$$

откуда

$$\frac{\partial \log \mathsf{L}(\lambda \mid \mathbf{x})}{\partial \lambda} = 0 \iff -n + \frac{n\bar{x}}{\lambda} = 0,\ n\bar{x} - n\lambda = 0,\ \lambda = \bar{x}.$$

Утверждение. В условиях регулярности:

1. Существует один глобальный максимум, так что

$$\left.\frac{\partial \log \mathsf{L}(\lambda \mid \mathbf{x})}{\partial \lambda}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\mathrm{MLE}}} = 0.$$

2. $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ обладает всеми свойствами (про определение этих свойст написано в следующих разделах):

   a) Состоятельность;

   b) Асимптотическая несмещенность;

   c) Асимптотическая нормальность;

   d) Асимптотическая эффективность.

[1] Maximum likelihood estimate (MLE).

6. Свойства оценок

6.1. Несмещенность

Определение. *Смещение*[1] есть

$$\operatorname{bias}\hat\theta_n := \mathsf{E}\hat\theta_n - \theta \quad \forall\theta\in\Theta.$$

Определение. *Среднеквадратичная ошибка*[2] есть

$$\operatorname{MSE}\hat\theta_n := \mathsf{E}(\hat\theta_n-\theta)^2.$$

Замечание. Поскольку

$$\mathsf{D}\hat\theta_n = \mathsf{D}(\hat\theta_n-\theta) = \mathsf{E}(\hat\theta_n-\theta)^2 - (\mathsf{E}(\hat\theta_n-\theta))^2,$$

то

$$\underbrace{\mathsf{E}(\hat\theta_n-\theta)^2}_{\text{MSE}} = \mathsf{D}\hat\theta_n + \underbrace{(\mathsf{E}(\hat\theta_n-\theta))^2}_{\text{bias}^2}.$$

Определение. Оценка называется *несмещенной*, если $\operatorname{bias}\hat\theta_n = 0$, т.е.

$$\mathsf{E}\hat\theta_n = \theta.$$

Предложение. *$\bar x$ — несмещенная оценка $\mathsf{E}\xi$.*

Доказательство. Пусть $\theta = \mathsf{E}\xi$, $\hat\theta_n = \widehat{\mathsf{E}\xi}_n = \bar x$. Тогда

$$\mathsf{E}\bar x = \mathsf{E}\frac{1}{n}\sum_{i=1}^n x_i = \frac{1}{n}\sum_{i=1}^n \mathsf{E}x_i = \frac{1}{n}\sum_{i=1}^n \mathsf{E}\xi = \mathsf{E}\xi \implies \mathsf{E}\hat\theta_n = \mathsf{E}\theta,\ \operatorname{bias}\hat\theta_n = 0.$$

□

Предложение. *s^2 является только* асимптотически *несмещенной оценкой $\mathsf{D}\xi$.*

Доказательство. Это доказательство 'в лоб'. Хорошее доказательство на основе разложения $\mathsf{E}(\zeta-a)^2$ смотрите в конспекте.

Поскольку дисперсия не зависит от сдвига, обозначим $\eta = \xi - \mathsf{E}\xi$ и $y_i = x_i - \mathsf{E}\xi$; тогда

$$\begin{aligned}
\mathsf{E}s^2 &= \mathsf{E}\widehat{\mathsf{D}\xi} = \mathsf{E}\widehat{\mathsf{D}\eta} = \mathsf{E}\left(\widehat{\mathsf{E}\eta^2} - \left(\widehat{\mathsf{E}\eta}\right)^2\right) = \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n y_i^2 - \left(\frac{1}{n}\sum_{i=1}^n y_i\right)^2\right)\\
&= \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n y_i^2 - \frac{1}{n^2}\sum_{i,j=1}^n y_iy_j\right) = \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n y_i^2 - \frac{1}{n^2}\sum_{i,j=1}^n y_i^2\right) = \frac{1}{n}\sum_{i=1}^n \mathsf{E}y_i^2 - \frac{1}{n^2}\sum_{i=1}^n \mathsf{E}y_i^2\\
&= \frac{1}{n}\sum_{i=1}^n \mathsf{E}\,(x_i-\mathsf{E}\xi)^2 - \frac{1}{n^2}\sum_{i=1}^n \mathsf{E}\,(x_i-\mathsf{E}\xi)^2 = \frac{1}{n}\sum_{i=1}^n \mathsf{D}x_i - \frac{1}{n^2}\sum_{i=1}^n \mathsf{D}x_i = \mathsf{D}\xi - \frac{1}{n}\mathsf{D}\xi\\
&= \frac{n-1}{n}\mathsf{D}\xi \xrightarrow[n\to\infty]{} \mathsf{D}\xi.
\end{aligned}$$

□

[1]Bias.
[2]Mean squared error (MSE).

Определение. *Исправленная* дисперсия:

$$\tilde{s}^2 := \frac{n}{n-1} s^2.$$

Очевидно, исправленная дисперсия — несмещенная оценка дисперсии.

6.2. Состоятельность

Определение. Оценка называется *состоятельной в среднеквадратичном смысле*, если

$$\operatorname{MSE} \hat{\theta}_n \xrightarrow[n\to\infty]{} 0.$$

Определение. Оценка называется *состоятельной*, если

$$\hat{\theta}_n \xrightarrow{\mathsf{P}} \theta.$$

Предложение. *Если оценка несмещенная и состоятельная в среднеквадратичном смысле, то она состоятельная.*

Доказательство. В самом деле, по неравенству Чебышева,

$$\mathsf{P}(|\hat{\theta}_n - \theta| > \epsilon) = \mathsf{P}(|\hat{\theta}_n - \mathsf{E}\hat{\theta}_n| > \epsilon) \le \frac{\mathsf{D}\hat{\theta}_n}{\epsilon^2} = \frac{\operatorname{MSE} \hat{\theta}_n}{\epsilon^2} \xrightarrow[n\to\infty]{} 0.$$

□

Предложение. $\hat{\mathsf{m}}_k$ *является состоятельной оценкой* m_k.

Доказательство. Докажем для $\hat{\mathsf{m}}_1$. По определению выборки до эксперимента, $x_i \sim \mathcal{P}$. Тогда, по теореме Хинчина о ЗБЧ,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \xrightarrow{\mathsf{P}} \mathsf{m}_1(\mathcal{P}).$$

Для k-го момента доказывается аналогично заменой $y_i := x_i^k$. □

Замечание. Для $\mathsf{m}_k^{(0)}$ доказательство не пройдет, потому что x_i и $\bar{x}$ не будут независимыми.

Предложение. $\hat{\mathsf{m}}_k^{(0)}$ *является состоятельной оценкой* $\mathsf{m}_k^{(0)}$.

Утверждение. Пусть $\xi_n \xrightarrow{\mathsf{P}} c$ и $f \in C(U_\epsilon(c))$. Тогда $f(\xi_n) \xrightarrow{\mathsf{P}} f(c)$.

Доказательство предложения. Докажем для s^2. Пусть $f : (x, y) \mapsto x - y^2$. Устроим последовательность $(\hat{\mathsf{m}}_2, \hat{\mathsf{m}}_1) \xrightarrow{\mathsf{P}} (\mathsf{m}_2, \mathsf{m}_1)$. Тогда

$$f(\hat{\mathsf{m}}_2, \hat{\mathsf{m}}_1) = \hat{\mathsf{m}}_2 - \hat{\mathsf{m}}_1^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2 = s^2 \xrightarrow{\mathsf{P}} f(\mathsf{m}_2, \mathsf{m}_1) = \mathsf{D}\xi.$$

Для $\mathsf{m}_k^{(0)}$ доказывается аналогично. □

Предложение. $\bar{x}$ *— состоятельная оценка* $\mathsf{E}\xi$.

Доказательство. Либо по (6.2) для $k = 1$, либо из того факта, что $\operatorname{bias} \bar{x} = 0$, значит

$$\operatorname{MSE} \bar{x} = \mathsf{D}\bar{x} = \frac{\mathsf{D}\xi}{n} \xrightarrow[n\to\infty]{} 0,$$

и по (6.2) получаем утверждение. □

Предложение. s^2 *— состоятельная оценка* $\mathsf{D}\xi$.

Доказательство. По (6.2) с $k = 2$. □

6.3. Асимптотическая нормальность

Определение. Оценка $\hat{\theta}_n$ называется *асимптотически нормальной* оценкой параметра θ с коэффициентом $\sigma^2(\theta)$ если

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2(\theta)).$$

Пример. $\bar{x}$ — асимптотически нормальная оценка, если $\mathsf{D}\xi < \infty,\ \mathsf{D}\xi \neq 0$:

$$\sqrt{n}\,(\bar{x} - \mathsf{E}\xi) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathsf{D}\xi).$$

Доказательство. По ЦПТ,

$$\sqrt{n}\,(\bar{x} - \mathsf{E}\xi) = \frac{\sum_{i=1}^n x_i - n\mathsf{E}\xi}{\sqrt{n}} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathsf{D}\xi).$$

□

Мы обсуждали, что асимптотическую нормальность можно определять и в слабом смысле — как сходимость по распределению к нормальному распределению $N(0,1)$ стандартизированной случайной величины.

6.4. Эффективность

Определение. Говорят, что оценка $\hat{\theta}^{(1)}$ *лучше* $\hat{\theta}^{(2)}$ *в среднеквадратичном смысле*, если

$$\mathrm{MSE}\,\hat{\theta}^{(1)} \leq \mathrm{MSE}\,\hat{\theta}^{(2)}.$$

Замечание. Для несмещенных оценок определение эквивалентно, конечно,

$$\mathsf{D}\hat{\theta}^{(1)} \leq \mathsf{D}\hat{\theta}^{(2)}.$$

Определение. В классе несмещенных оценок оценка называется эффективной (в средне-квадратическом), если ее дисперсия минимальна. В классе асимптотически несмещенных оценок оценка $\hat{\theta}$ называется асимптотически эффективной, если для любой другой оценки $\hat{\theta}^*$ выполнено $\mathsf{D}\hat{\theta} \leq \mathsf{D}\hat{\theta}^* \leq 1$.

6.4.1. Эффективность и неравенство Рао-Крамера

Пусть $\mathcal{P}_\xi(\boldsymbol{\theta}),\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)^\mathsf{T}$ — параметрическая модель. Пусть $r = 1$.

Определение. *Информанта n-го порядка*:

$$S_n(\mathbf{x}, \theta) = \frac{\mathrm{d}^n \ln \mathsf{L}(\theta \mid \mathbf{x})}{\mathrm{d}\theta^n}.$$

Определение. *Информационное количество Фишера*:

$$I_n(\theta) := -\mathsf{E}S_2(\mathbf{x}, \theta).$$

Утверждение.

$$I_n(\theta) = \mathsf{E}S_1^2(\mathbf{x}, \theta).$$

Пример. $\xi \sim \mathrm{Pois}(\lambda)$.

$$S_1(\mathbf{x}, \theta) = -n + \frac{n\bar{x}}{\lambda}, \quad S_2(\mathbf{x}, \theta) = -\frac{n\bar{x}}{\lambda^2} \implies I_n(\lambda) = \mathsf{E}\frac{n\bar{x}}{\lambda^2} = \frac{n}{\lambda^2}\mathsf{E}\bar{x} = \frac{n}{\lambda}.$$

Замечание.

$$\ln \mathsf{L}(\theta \mid \mathbf{x}) = \sum_{i=1}^{n} \ln p_\theta(x_i) \implies S_2 = \frac{\mathrm{d}^2 \ln \mathsf{L}(\theta \mid \mathbf{x})}{\mathrm{d}\theta^2} = \sum_{i=1}^{n} (\ln p_\theta(x_i))'',$$

откуда, для повторной независимой выборки,

$$I_n(\theta) = -\sum_{i=1}^{n} \mathsf{E}(\ln p_\theta(x_i))'' = n \cdot i(\theta), \quad \text{где } i(\theta) = -\mathsf{E}(\ln p_\theta(\xi))''.$$

Определение. $C \subset \mathbb{R}$ есть *носитель* параметрического семейства распределений $\mathcal{P}(\theta)$, если

$$\xi \sim \mathcal{P}(\theta) \implies \mathsf{P}(\xi \in C) = 1, \ \forall \theta \in \Theta.$$

Определение. Условие регулярности: имеют отношение к независимости носителя распределения от параметра, а также к существованию и ограниченности производных функции логправдоподобия по параметру до определённого порядка дифференцирования.

Пример. $\mathrm{Exp}(\lambda)$ — регулярное семейство; $\mathrm{U}(0, \theta)$ — не является регулярным.

Утверждение. Для несмещенных оценок в условиях регулярности справедливо неравенство *Рао–Крамера*:

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

Для смещенных оценок,

$$\mathsf{D}\hat{\theta}_n \geq \frac{(1 + \mathrm{bias}'(\theta))^2}{I_n(\theta)}.$$

Следствие. *Несмещенная оценка является эффективной, если:*

$$\mathsf{D}\hat{\theta}_n = \frac{1}{I_n(\theta)}.$$

Следствие. *Асимптотически несмещенная оценка является асимптотически эффективной, если:*

$$\mathsf{D}\hat{\theta}_n \cdot I_n(\theta) \to 1 \ \textit{as} \ n \to \infty.$$

Упражнение (Хорошее)**.** Показать, что $\bar{x}$ является эффективной оценкой μ в модели $\xi \sim \mathrm{N}(\mu, \sigma^2)$.

Следствие. *Пусть $\hat{\theta}_n$ — асимптотически несмещенная оценка. Тогда $\hat{\theta}_n$ —* асимптотически эффективная*, если*

$$\mathsf{D}\hat{\theta}_n \cdot I_n \xrightarrow[n \to \infty]{} 1.$$

Пример. Пусть $\xi \sim \mathrm{N}(\mu, \sigma^2)$. Можно посчитать, что s^2 является только асимптотически эффективной оценкой σ^2; $\tilde{s}^2$ — просто эффективной.

Пример. Пусть $\xi \sim \mathrm{Pois}(\lambda)$. Поскольку

$$\begin{aligned} \mathsf{D}\hat{\lambda}_n &= \mathsf{D}\bar{x} = \mathsf{E}\xi/n = \lambda/n \\ I_n(\lambda) &= n/\lambda, \end{aligned}$$

то $\hat{\lambda}_n$ — эффективная оценка (как и ожидалась по свойствам $\hat{\theta}_{\mathrm{MLE}}$).

6.5. Устойчивость оценок

Так как в реальных данных часто бывают те или иные ошибки, часто жертвуют точностью для увеличения устойчивости (робастности) к выбросам. Устойчивые аналоги оценок часто строятся на основе рангов (номеров по порядку в упорядоченной выборке). Приведем пример.

Пример (Сравнение оценок мат. ожидания симметричного распределения)**.** Пусть $\mathcal{P}$ *симметрично* — в этом случае $\widehat{\operatorname{med}\xi} = \bar{x}$ и имеет смысл сравнить две этих характеристики.

$$\begin{aligned} \mathsf{D}\bar{x} &= \frac{\mathsf{D}\xi}{n} \\ \mathsf{D}\widehat{\operatorname{med}\xi} &\sim \frac{1}{4n\operatorname{pdf}^2_{\mathrm{N}(\mu,\sigma^2)}(\operatorname{med}\xi)} \quad \text{при } n \to \infty. \end{aligned}$$

Так, если $\xi \sim \mathrm{N}(\mu, \sigma^2)$, то

$$\operatorname{pdf}^2_{\mathrm{N}(\mu,\sigma^2)}(\operatorname{med}\xi) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{(\operatorname{med}\xi - \mu)^2}{\sigma^2}\right\} = \frac{1}{2\pi\sigma^2},$$

откуда

$$\mathsf{D}\widehat{\operatorname{med}\xi} = \frac{\pi}{2}\frac{\sigma^2}{n} > \frac{\sigma^2}{n} = \mathsf{D}\bar{x},$$

значит $\bar{x}$ эффективнее $\widehat{\operatorname{med}\xi}$.

Замечание. В то же время, $\widehat{\operatorname{med}\xi}$ более устойчива к аутлаерам, чем $\bar{x}$, и этим лучше.

Часть II.

Некоторые распределения, связанные с нормальным

1. Распределение $N(a, \sigma^2)$

Свойства хорошо известны. TODO

Здесь пока напишу только про измерение расстояния в сигмах. Будет говорить, что точка далеко от мат.ожидания, если это и более далекие значения маловероятны.

Формально, пусть $\xi \sim N(a, \sigma^2)$. Рассмотрим $\mathsf{P}(|\xi - a| > k\sigma)$. Эта вероятность не зависит от σ и равна $2(1 - \Phi(k))$, где $\Phi(x)$ — функция стандартного нормального распределения.

Значения $\mathsf{P}(|\xi - a| > k\sigma)$:

k	Вероятность
1	0.317
1.64	0.101
1.96	0.050
2	0.046
3	2.70E-03
6	1.97E-09

2. Распределение $\chi^2(m)$

Определение (Распределение $\chi^2(m)$)**.** η имеет распределение χ^2 с m степенями свободны:

$$\eta \sim \chi^2(m) \iff \eta = \sum_{i=1}^{m} \zeta_i^2, \quad \zeta_i \sim \mathrm{N}(0,1),\ \zeta_i \text{ независимы.}$$

Свойства[1] $\chi^2(m)$

$$\begin{aligned} \mathsf{E}\eta &= \sum_{i=1}^{m} \mathsf{E}\zeta_i^2 = m \\ \mathsf{D}\eta &= 2m \end{aligned}$$

Утверждение. Пусть $\eta_m \sim \chi^2(m)$. Тогда, по ЦПТ,

$$\frac{\eta_m - \mathsf{E}\eta_m}{\sqrt{\mathsf{D}\eta_m}} = \frac{\eta_m - m}{\sqrt{2m}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Пример. $m = 50,\ \eta_m = 80$. Тогда

$$\frac{80-50}{10} = 3$$

и

$$\mathrm{cdf}_{\chi^2(50)}(80) = 0.9955 \approx \Phi(3) = 0.9986.$$

Предложение. $\chi^2(m)/m \xrightarrow[m\to\infty]{} 1$.

Доказательство. По ЗБЧ. □

[1]Вычисление $\mathsf{D}\eta$: https://www.statlect.com/probability-distributions/chi-square-distribution

3. Распределение Стьюдента $\mathrm{t}(m)$

Определение (Распределение $\mathrm{t}(m)$)**.** ξ имеет распределение Стьюдента с m степенями свободы, если

$$\xi \sim \mathrm{t}(m) \iff \xi = \frac{\zeta}{\sqrt{\eta/m}}, \quad \zeta \sim \mathrm{N}(0,1),\ \eta \sim \chi^2(m).$$

Свойства $\mathrm{t}(m)$

- При $m = 1$ это распределение Коши.
- При $m > 1$, $\mathsf{E}\xi = 0$ по симметричности.
- При $m > 2$, $\mathsf{D}\xi = m/(m-2)$.
- При $m > 3$, $\mathsf{A}\xi = 0$ по симметричности.
- При $m > 4$, $\mathsf{K}\xi = 6/(m-4)$.

Предложение. *Распределение Стьюдента сходится к стандартному нормальному:*

$$\mathrm{t} \Rightarrow \mathrm{N}(0,1).$$

Соображения по поводу. $\mathsf{D}\xi \to 1,\ \mathsf{K}\xi \to 0.$ $\square$

4. Распределение Фишера

Определение. Распределение Фишера имеет вид

$$\mathrm{F}(m,k) = \frac{\chi^2(m)/m}{\chi^2(k)/k}.$$

Замечание. $\mathrm{F}(1,k) \sim \mathrm{t}^2(k)$; $\mathrm{F}(m,\infty) = \chi^2(m)/m$ потому что $\chi^2(k)/k \xrightarrow[k\to\infty]{} 1$.

5. Квадратичные формы от нормально распределенных случайных величин

(Это на след. семестр, сейчас можно не вникать.)
Пусть $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^\mathsf{T} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$, $\mathbf{B}$ — симметричная, неотрицательно определенная матрица. Найдем распределение $\boldsymbol{\xi}^\mathsf{T}\mathbf{B}\boldsymbol{\xi}$.

Утверждение. Пусть $\boldsymbol{\xi} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$, $\mathbf{B}, \mathbf{C}$ — симметричные матрицы размерности $p \times p$. Тогда $\boldsymbol{\xi}^\mathsf{T}\mathbf{B}\boldsymbol{\xi} \perp\!\!\!\perp \boldsymbol{\xi}^\mathsf{T}\mathbf{C}\boldsymbol{\xi} \iff \mathbf{BC} = \mathbf{0}$.

Пример (Независимость $\bar{x}^2$ и s^2)**.** Запишем

$$\begin{aligned}
\bar{x}^2 &= \frac{1}{n^2}\left(\sum_{i=1}^n x_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n x_i x_j = \frac{1}{n}\mathbf{x}\underbrace{\begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \dots & 1/n \end{pmatrix}}_{\mathbf{B}}\mathbf{x}^\mathsf{T} \\
s^2 &= \frac{1}{n}\sum_{i=1}^n x_i^2 - \frac{1}{n}\mathbf{x}\mathbf{B}\mathbf{x}^\mathsf{T} = \frac{1}{n}\left(\mathbf{x}\mathbf{I}_n\mathbf{x}^\mathsf{T} - \mathbf{x}\mathbf{B}\mathbf{x}^\mathsf{T}\right) = \frac{1}{n}\mathbf{x}\underbrace{\begin{pmatrix} 1-1/n & \dots & -1/n \\ \vdots & \ddots & \vdots \\ -1/n & \dots & 1-1/n \end{pmatrix}}_{\mathbf{C}=\mathbf{I}_n-\mathbf{B}}\mathbf{x}^\mathsf{T}.
\end{aligned}$$

Таким образом, $n\bar{x}^2 = \mathbf{x}\mathbf{B}\mathbf{x}^\mathsf{T}$ и $ns^2 = \mathbf{x}\mathbf{C}\mathbf{x}^\mathsf{T}$. Но

$$\mathbf{BC} = \mathbf{B}(\mathbf{I}_n - \mathbf{B}) = \mathbf{B} - \mathbf{B}^2 = \mathbf{0},$$

так как

$$\mathbf{B}^2 = \begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \dots & 1/n \end{pmatrix}^2 = \begin{pmatrix} n \cdot 1/n & \dots & n \cdot 1/n \\ \vdots & \ddots & \vdots \\ n \cdot 1/n & \dots & n \cdot 1/n \end{pmatrix} = \mathbf{B}.$$

Значит, $\bar{x}^2 \perp\!\!\!\perp s^2$.

Видно, что $\sigma^{-2}\boldsymbol{\xi}^\mathsf{T}\mathbf{I}_p\boldsymbol{\xi} \sim \chi^2(p)$. На самом деле справедливо

Утверждение. Пусть $\boldsymbol{\xi} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$, $\mathbf{B}$ — симметричная, неотрицательно неопределенная матрица размерности $p \times p$ и $\operatorname{rk}\mathbf{B} = r$. Тогда

$$\sigma^{-2}\boldsymbol{\xi}^\mathsf{T}\mathbf{B}\boldsymbol{\xi} \sim \chi^2(r) \iff \mathbf{B}^2 = \mathbf{B}.$$

Пример. Покажем, что

$$n\sigma^{-2}s^2 \sim \chi^2(p-1).$$

Воспользуемся представлением из предыдущего примера: $ps^2 = \mathbf{x}^\mathsf{T}\mathbf{C}\mathbf{x}$. Но $\operatorname{rk}\mathbf{C} = \operatorname{rk}(\mathbf{I}_p - \mathbf{B}) = p - 1$; $\mathbf{B}^2 = \mathbf{B}$, значит $p\sigma^{-2}s^2 \sim \chi^2(p-1)$.

Утверждение (Cochran). Пусть $\boldsymbol{\xi} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_p)$, $\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\xi} = \sum_i Q_i$, где Q_i — квадратичная форма, заданная $\mathbf{B}_i$, $\operatorname{rk}\mathbf{B}_i = r_i$. Тогда следующие утверждения эквивалентны:

1. $\sum r_i = p$
2. $Q_i \sim \chi^2(r_i)$
3. $Q_i \perp\!\!\!\perp Q_j, \quad \forall i \neq j$, т.е. $\mathbf{B}_i\mathbf{B}_j = \mathbf{0}$.

6. Распределение важных статистик

Пусть $\xi \sim \mathrm{N}(a, \sigma^2)$.

Предложение. $t = \sqrt{n}\frac{(\bar{x} - \mathsf{E}\xi)}{\sigma}$ *имеет стандартное нормальное распределение.*

Доказательство.

$$t = \frac{\bar{x} - a}{\sqrt{\mathsf{D}\bar{x}}} = \sqrt{n}\,\frac{\bar{x} - a}{\sigma} \sim \mathrm{N}(0, 1).$$

□

Определим $s_a^2 = \sum_{i=1}^n (x_i - a)^2/n$.

Предложение. $ns_a^2/\sigma^2 \sim \chi^2(n)$.

Доказательство.

$$\chi^2 = \frac{ns_a^2}{\sigma^2} = \frac{n \cdot 1/n \cdot \sum_{i=1}^n (x_i - a)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - a}{\sigma}\right)^2 \sim \chi^2(n).$$

□

Предложение. $ns^2/\sigma^2 = (n-1)\tilde{s}^2/\sigma^2 \sim \chi^2(n-1)$.

Доказательство. См. раздел 5). □

Альтернативное доказательство. По определению запишем

$$\underbrace{\mathsf{D}\hat{\xi}_n}_{s^2} = \mathsf{D}(\hat{\xi}_n - a) = \underbrace{\mathsf{E}(\hat{\xi}_n - a)^2}_{s_a^2} - \underbrace{(\mathsf{E}(\hat{\xi}_n - a))^2}_{(\bar{x}-a)^2}.$$

Домножив обе части на n/σ^2, получим

$$\frac{ns^2}{\sigma^2} = \frac{ns_a^2}{\sigma^2} - \frac{n(\bar{x} - a)^2}{\sigma^2} = \underbrace{\frac{ns_a^2}{\sigma^2}}_{\sim\chi^2(n)} - \underbrace{\left(\frac{\sqrt{n}\,(\bar{x} - a)}{\sigma}\right)^2}_{\sim\chi^2(1)} \implies \frac{ns^2}{\sigma^2} \sim \chi^2(n-1).$$

□

Замечание. Для строгого доказательства, нужно использовать независимость $\bar{x}^2$ и s^2 (см. раздел 5).

Предложение. *Следующая статистика имеет распределение Стьюдента:*

$$t = \sqrt{n-1}\,\frac{\bar{x} - a}{s} = \frac{\sqrt{n-1}(\bar{x} - a)}{\sqrt{n-1}/\sqrt{n} \cdot \tilde{s}} = \sqrt{n}\,\frac{\bar{x} - a}{\tilde{s}} \xrightarrow[n\to\infty]{} \mathrm{N}(0, 1).$$

Предложение. $t = \sqrt{n-1}\,\frac{\bar{x}-a}{s} = \sqrt{n}\,\frac{\bar{x}-a}{\tilde{s}} \sim \mathrm{t}(n-1)$.

Доказательство.

$$t = \frac{\sqrt{n-1}\,(\bar{x}-a)}{s} = \frac{\sqrt{n-1}\left(\dfrac{\bar{x}-a}{\sigma}\right)}{s/\sigma} = \frac{\left(\dfrac{\bar{x}-a}{\sigma}\right)}{\sqrt{\dfrac{s^2/\sigma^2}{n-1}}} = \frac{\dfrac{\sqrt{n}(\bar{x}-a)}{\sigma}}{\sqrt{\dfrac{ns^2/\sigma^2}{n-1}}} = \frac{\zeta}{\sqrt{\eta/(n-1)}} \sim \mathrm{t}(n-1),$$

поскольку

$$\zeta = \frac{\sqrt{n}(\bar{x}-a)}{\sigma} \sim \mathrm{N}(0,1), \quad \eta = \frac{ns^2}{\sigma^2} \sim \chi^2(n-1).$$

и они независимы (также пока без доказательства — используется техника квадратичных форм или можно доказать через разложение дисперсии). □

Часть III.

Проверка гипотез и доверительные интервалы

Этот раздел иногда называется «Confirmatory Data Analysis» в противовес «Exploratory Data Analysis», не включающему в себя понятие *гипотезы*.

1. Построение критерия

Тут должно быть про определение критерия и построение его через статистику критерия.

2. Проверка гипотезы о значении параметра (характеристики)

2.1. Проверка гипотезы о значении мат. ожидания (t-критерий)

$H_0 : \mathsf{E}\xi = a = a_0$. Соответствие оценки математического ожидания гипотезе удобно выражать разницей $\bar{x} - a_0$ с «идеальным» значением 0. Отнормировав эту разницу, получим статистику, распределение которой известно.

2.1.1. $\mathsf{D}\xi = \sigma^2 < \infty$

Предложение. *Пусть $\mathsf{D}\xi = \sigma^2 < \infty$; тогда используется следующая статистика (z-score)*

$$t = z = \sqrt{n}\frac{(\bar{x} - a_0)}{\sigma} \xrightarrow[n\to\infty]{} \mathrm{N}(0,1)$$

Предложение. *При условии $\xi \sim \mathrm{N}(a, \sigma^2)$,*

$$t = z \sim \mathrm{N}(0,1).$$

Доказательство.

$$z = \frac{\bar{x} - a_0}{\sqrt{\mathsf{D}\bar{x}}} = \sqrt{n}\,\frac{\bar{x} - a_0}{\sigma} \sim \mathrm{N}(0,1).$$

□

2.1.2. $\mathsf{D}\xi$ неизвестна

Предложение. *Пусть $\mathsf{D}\xi$ неизвестна; тогда используется следующая статистика*

$$t = \sqrt{n-1}\,\frac{\bar{x} - a_0}{s} = \sqrt{n}\,\frac{\bar{x} - a_0}{\tilde{s}} \xrightarrow[n\to\infty]{} \mathrm{N}(0,1).$$

Предложение. *При условии нормальности данных,*

$$t \sim \mathrm{t}(n-1).$$

2.1.3. z-критерий для пропорции в модели Бернулли

Пусть $\xi \sim \mathrm{Ber}(p)$. Поскольку $\mathsf{E}\xi = p$, можно воспользоваться только что введенной статистикой; учитывая $\mathsf{D}\xi = p(1-p)$,

$$T = \sqrt{n}\frac{\bar{x} - p_0}{\sqrt{p_0(1-p_0)}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Разница с общим случаем состоит в том, что в параметрической модели с одним параметром не нужно оценивать дисперсию. Так как все выражается через этот параметр, то имеет формулу для дисперсии через значение параметра, предполагаемое в нулевой гипотезе.

2.2. Проверка гипотезы о значении дисперсии в нормальной модели (критерий χ^2)

Пусть $\xi \sim \mathrm{N}(a, \sigma^2)$. $H_0 : \mathsf{D}\xi = \sigma^2 = \sigma_0^2$. Соответствие оценки дисперсии гипотезе удобно выражать отношением s^2/σ_0^2 (или s_a/σ_0^2 если a известно) с «идеальным» значением 1. Домножив на n, получим статистику, распределение которой известно.

2.2.1. $\mathsf{E}\xi = a < \infty$

Предложение. *Пусть* $\mathsf{E}\xi = a < \infty$*; При условии нормальности данных используется следующая статистика:*

$$\chi^2 = n\frac{s_a^2}{\sigma_0^2} \sim \chi^2(n).$$

2.2.2. $\mathsf{E}\xi$ неизвестно

Предложение. *Пусть* $\mathsf{E}\xi$ *неизвестно. При условии нормальности данных используется следующая статистика:*

$$\chi^2 = n\frac{s^2}{\sigma_0^2} = (n-1)\frac{\tilde{s}^2}{\sigma_0^2} \sim \chi^2(n-1).$$

Упражнение. $s^2 = 1.44, \bar{x} = 55, n = 101$. Проверить гипотезу $\sigma_0^2 = 1.5$ в нормальной модели.

Решение. Воспользуемся статистикой

$$\chi^2 = \frac{ns^2}{\sigma_0^2} = 101 \cdot 0.96 = 96.96.$$

«Идеальные» значения близки к $\mathsf{E}\xi_{\chi^2(100)} = 100$, так что определим критическую область на концах плотности:

$$p\text{-value}/2 = \mathrm{cdf}_{\chi^2(100)}(96.96) = \texttt{pchisq}(96.96, 100) \approx 0.43 \implies p\text{-value} \approx 0.86.$$

Замечание. Можно посчитать и по таблицам для нормального распределения. Раз

$$\frac{\eta_m - \mathsf{E}\eta_m}{\sqrt{\mathsf{D}\eta_m}} \xrightarrow[m\to\infty]{\mathrm{d}} \mathrm{N}(0,1),$$

то

$$\frac{96.96 - 100}{\sqrt{200}} \approx -0.215 \implies p\text{-value}/2 = \Phi(-0.215) \approx 0.415.$$

⌟

3. Доверительные интервалы

3.1. Мотивация и определение

Точечные оценки не дают информации о том, насколько (количественно) настоящее значение далеко от оценки.

Определение. $[b_1, b_2]$ — *доверительный интервал* для параметра θ с уровнем доверия $\gamma \in [0,1]$, если $\forall\theta$

$$\mathsf{P}(\theta \in [b_1, b_2]) = \gamma, \quad \text{где } b_1 = b_1(\mathbf{x}), b_2 = b_2(\mathbf{x}) \text{ — статистики.}$$

Замечание. Если выборка из дискретного распределения, то b_1, b_2 — тоже дискретны. Поэтому наперед заданную точность получить может не получиться; в таких случаях знак «=» заменяют «$\geq$». Аналогично с заменой на «$\xrightarrow[n\to\infty]{}$» для асимптотических доверительных интервалов, когда точные получить невозможно или трудоемко.

3.2. Доверительный интервал для проверки гипотезы о значении параметра

Зафиксируем $H_0 : \theta = \theta_0$ и $\gamma = 1 - \alpha$, где α играет роль уровня значимости. По определению доверительного интервала, $\mathsf{P}(\theta \in [a_\gamma(\mathbf{x}), b_\gamma(\mathbf{x})]) = \gamma$. Тогда

$$\mathsf{P}(\theta \in [b_1(\mathbf{x}), b_2(\mathbf{x})]) = \gamma = 1 - \alpha \implies \alpha = 1 - \mathsf{P}(\theta \in [a_\gamma(\mathbf{x}), b_\gamma(\mathbf{x})]) = \mathsf{P}(\theta \notin [a_\gamma(\mathbf{x}), b_\gamma(\mathbf{x})])$$

и $\mathscr{A}_{\text{крит}} = \mathbb{R} \setminus [b_1(\mathbf{x}), b_2(\mathbf{x})]$. Соответственно,

$$\begin{cases} \text{отвергаем } H_0\text{, если} & \theta_0 \notin [b_1(\mathbf{x}), b_2(\mathbf{x})] \\ \text{не отвергаем } H_0\text{, если} & \theta_0 \in [b_1(\mathbf{x}), b_2(\mathbf{x})] \,. \end{cases}$$

Вероятность ошибки первого рода равно α, что соответствует определению критерия. Заметим, что здесь мы пользуемся общим определением критерия, а не частным случаем, когда критерий строится через статистику критерия.

3.3. Доверительные интервалы для математического ожидания и дисперсии в нормальной модели

Предположение. *Пусть* $\xi \sim \mathrm{N}(a, \sigma^2)$.

3.3.1. Доверительный интервал для a

- Пусть σ^2 известно. Свяжем a с выборкой через статистику критерия $t = \sqrt{n}\frac{(\bar{x}-a)}{\sigma} \sim \mathrm{N}(0,1)$:

  $$\gamma = \mathsf{P}(c_1 < t < c_2) = \mathsf{P}\left(c_1 < \sqrt{n}\frac{(\bar{x}-a)}{\sigma} < c_2\right) = \mathsf{P}\left(a \in \left(\bar{x} - \frac{\sigma c_2}{\sqrt{n}}, \bar{x} - \frac{\sigma c_1}{\sqrt{n}}\right)\right).$$

  Решений уравнения $\mathsf{P}\left(c_1 < \sqrt{n}(\bar{x}-a)/\sigma < c_2\right) = \Phi(c_2) - \Phi(c_1) = \gamma$ бесконечно много. Чем $[c_1, c_2]$ короче, тем лучше. Поскольку Φ симметрична и унимодальна,

  $$\begin{aligned} c_1 &= -c_\gamma \\ c_2 &= c_\gamma, \end{aligned} \quad \text{где } c_\gamma = \mathrm{cdf}^{-1}_{\mathrm{N}(0,1)}\left(\gamma + \frac{1-\gamma}{2}\right) = x_{\frac{1+\gamma}{2}}.$$

Наконец,

$$\mathsf{P}\left(a \in \left(\bar{x} \pm \frac{\sigma}{\sqrt{n}} c_\gamma\right)\right) = \gamma.$$

- Пусть σ^2 неизвестно. По аналогии,

$$\gamma = \mathsf{P}\left(c_1 < \frac{\sqrt{n-1}(\bar{x}-a)}{s} < c_2\right) = \mathsf{P}\left(a \in \left(\bar{x} \pm \frac{c_\gamma s}{\sqrt{n-1}}\right)\right), \quad c_\gamma = \mathrm{cdf}^{-1}_{\mathrm{t}(n-1)}\left(\frac{1+\gamma}{2}\right)$$

и

$$\mathsf{P}\left(a \in \left(\bar{x} \pm \frac{\tilde{s}}{\sqrt{n}} c_\gamma\right)\right) = \gamma.$$

Упражнение. Пусть $s^2 = 1.21, \bar{x} = 1.9, n = 36$. Построить 95% доверительный интервал для $\mathsf{E}\xi$.

Решение.

$$c_\gamma = \mathtt{qt}(0.975, 35) \approx 2.03 \implies \left(1.9 \pm \frac{2.03 \cdot \sqrt{1.21}}{\sqrt{35}}\right) = (1.52; 2.28).$$

┘

3.3.2. Доверительный интервал для σ^2

- Пусть a известно. Поскольку плотность χ^2 становится все более симметричной с ростом n, примем

$$c_1 = \mathrm{cdf}^{-1}_{\chi^2(n)}\left(\frac{1-\gamma}{2}\right), \ c_2 = \mathrm{cdf}^{-1}_{\chi^2(n)}\left(\frac{1+\gamma}{2}\right).$$

Тогда

$$\mathsf{P}\left(c_1 < \frac{ns_a^2}{\sigma^2} < c_2\right) = \gamma \iff \mathsf{P}\left(\sigma^2 \in \left(\frac{ns_a^2}{c_2}, \frac{ns_a^2}{c_1}\right)\right) = \gamma.$$

- Пусть a неизвестно. Тогда аналогично

$$\mathsf{P}\left(\sigma^2 \in \left(\frac{ns^2}{c_2}, \frac{ns^2}{c_1}\right)\right) = \gamma,$$

где

$$c_1 = \mathrm{cdf}^{-1}_{\chi^2(n-1)}\left(\frac{1-\gamma}{2}\right), \ c_2 = \mathrm{cdf}^{-1}_{\chi^2(n-1)}\left(\frac{1+\gamma}{2}\right).$$