

Математическая статистика

Матвеев Сергей М3338

5 семестр

1 Введение

Пусть у нас есть генеральная совокупность, но мы хотим ее как-то изучать, тогда мы можем взять ее часть - выборку.

Мы хотим по выборке сделать содержательные вероятностные выводы о генеральной совокупности.

Примеры задач, которые могут быть решены таким способом:

1. Бросок монеты (оценить вероятность орла, честно|нечестно)
2. Замеры показателя: какие типичные значения для показателя
3. Как учатся мальчики и девочки (одинаково или по разному)
4. Цена на недвижимость, расстояние до метро (оценка зависимости)

Репрезентативность: на основе выборки можно сделать выводы о генеративной совокупности.

2 Простейшая модель выборки. Эмпирическая функция распределения.

Простейшая модель выборки - X_1, X_2, \dots, X_n - *i.i.d.*, F - функция распределения (теоретическая функция).

$X_1, \dots, X_n \sim F$ (F мы не знаем априори)

x_1, \dots, x_n - реализация выборки

Цель: оценить из реализации x_1, \dots, x_n теор F .

Эмпирическая функция распределения:

$$\mu_n(t) = \sum_{i=1}^n \mathbb{1}(X_i \leq t)$$

$$F_n(t) = \frac{\mu_n(t)}{n} \text{ - эмпирическая функция распределения.}$$

Замечание: Описывает эмпирическое распределению.

$$x_1, \dots, x_n; P(U = x_i = \frac{\#\{X_j : X_j = x_i\}}{n} = F_n(x_i + 0) - F_n(x_i)$$

$$\mathbb{1}(\mathbb{X}_i \leq \mathbb{t}) \sim \text{Bern}(F(t))$$

$$\mathbb{E}(F_n(t)) = F(t) \text{ (это называется несмещенность)}$$

$$\text{Var}(F_n(t)) = \frac{F(t)(1 - F(t))}{n}$$

ЗБЧ: $F_n(t) \xrightarrow{P} F(t)$ - это называется состоятельность

$$\text{ЦПТ: } \frac{\mu_n(t) - nF(t)}{\sqrt{F(t)(1 - F(t))n}} \xrightarrow{d} U \sim N(0, 1) =$$

$$= \sqrt{n} \frac{F_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \text{ (асимптотическая нормальность)}$$

Теорема Гливенко-Кантелли

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$$

Теорема Колмогорова

$$D_n = \sup_x |F_n(x) - F(x)| \Rightarrow P(\sqrt{n}D_n \leq t) \xrightarrow[n \rightarrow \infty]{} K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 t^2}$$

$$F \in C(\mathbb{R})$$

Такая функция называется функцией Колмогорова

Теорема Смирнова

$X_1, \dots, X_n, Y_1, \dots, Y_n$ - независимы

Обе распределены по $F \in C(\mathbb{R})$

$$D_{n,m} = \sup_x |F_n(x) - F_m(x)| \Rightarrow P(\sqrt{\frac{mn}{m+n}} D_{n,m} \leq t) \xrightarrow[n \rightarrow \infty, m \rightarrow \infty]{} K(t)$$

Стоит отметить, что обе теоремы имеют смысл при $t \geq 0$

3 Выборочные моменты

$\alpha_k = EX_1^k$ - k -ый теоретический момент.

$\beta_k = E(X_1 - EX_1)^k$ - k -ый центральный момент.

$$\overline{g(X)} = \frac{1}{n} \sum_{k=1}^n g(X_k), g : \mathbb{R} \rightarrow \mathbb{R}$$

$$\widehat{\alpha}_k = \overline{X^k} = \frac{1}{n} \sum_{j=1}^n X_j^k \text{ - } k\text{-ый выборочный момент.}$$

$E\widehat{\alpha}_k = \alpha_k$ (несмещенность, мы просто воспользовались линейностью математического ожидания)

$$\text{Var } \widehat{\alpha}_k = \frac{1}{n} \text{Var}(X_1^k) = \frac{1}{n} (EX_1^{2k} - (EX_1^k)^2)$$

По ЦПТ получаем:

$$\sqrt{n} \frac{\widehat{\alpha}_k - \alpha_k}{\alpha_{2k} - \alpha_k^2} \approx N(0, 1)$$

$\sqrt{n} \frac{\widehat{\alpha}_k - \alpha_k}{\widehat{\alpha}_{2k} - \widehat{\alpha}_k^2} = \sqrt{n} \frac{\widehat{\alpha}_k - \alpha_k}{\alpha_{2k} - \alpha_k^2} \cdot \frac{\alpha_{2k} - \alpha_k^2}{\widehat{\alpha}_{2k} - \widehat{\alpha}_k^2}$ - первый множитель по ЦПТ сходится к $N(0, 1)$ по распределению

Давайте посмотрим что будет со вторым множителем. Он будет сходиться к 1 по вероятности.

Таким образом:

$$\sqrt{n} \frac{\widehat{\alpha}_k - \alpha_k}{\alpha_{2k} - \alpha_k^2} \cdot \frac{\alpha_{2k} - \alpha_k^2}{\widehat{\alpha}_{2k} - \widehat{\alpha}_k^2} \xrightarrow{d} N(0, 1)$$

А почему вторая дробь сходится к единице?

$\widehat{\alpha}_k \xrightarrow{P} \alpha_k$ (по ЗБЦ, это называется состоятельность)

$$\widehat{\alpha}_{2k} - \widehat{\alpha}_k^2 \xrightarrow{P} \alpha_{2k} - \alpha_k^2$$

\bar{X} - выборочное среднее.

$$\widehat{\beta}_k = \overline{(X - \bar{X})^k} = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^k - k\text{-ый выборочный момент.}$$

$\widehat{\beta}_2 = S_*^2$ - выборочная дисперсия.

S_* - выборочное стандартное отклонение (выборочное среднеквадратичное отклонение).

Note : выборочные моменты есть ничто иное как моменты посчитанные относительно эмпирического распределения.

$$S_*^2 = \overline{X^2} - (\bar{X})^2$$

$$\widehat{\beta}_k = Poly(\widehat{\alpha}_k, \dots, \widehat{\alpha}_1)$$

$\widehat{\alpha}_1, \dots, \widehat{\alpha}_k$ - состоятельные оценки (имеет место сходимость по вероятности)

Так как полином это непрерывная функция, то $\widehat{\beta}_k \xrightarrow{P} \beta_k$

TODO

4 Небольшое отступление

Пусть ξ_n - последовательность случайных векторов.

$$\sqrt{n}(\xi_n - \mu) \xrightarrow{P} N(0, \Sigma)$$

$\xi_n \xrightarrow{P} \mu$??? Давайте убедимся в этом

$$(\xi_n - \mu)\sqrt{n} \cdot \frac{1}{\sqrt{n}} \xrightarrow{d} 0 \text{ (Первое сходится к нормальному многомерному закону,}$$

а второе к нулю)

Мы знаем, что для вырожденных величин сходимость по распределению и вероятности равносильны.

$$\text{Тогда мы можем написать } (\xi_n - \mu)\sqrt{n} \cdot \frac{1}{\sqrt{n}} \xrightarrow{P} 0$$

$$\sqrt{n}(\varphi(\xi_n) - \varphi(\mu)) \rightarrow ??$$

$$\varphi \in C^1(\mathbb{R}^m), \varphi : \mathbb{R}^m \rightarrow \mathbb{R}$$

По Тейлору:

$$\varphi(\xi_n) \approx \varphi(\mu) + \nabla \varphi(\tilde{\mu})(\xi_n - \mu) \text{ (Остаток в форме Лагранжа)}$$

$$\nabla \varphi(\tilde{\mu}) \xrightarrow{n \rightarrow \infty} \nabla \varphi(\mu)$$

$$\varphi(\xi_n) - \varphi(\mu) \approx \nabla \varphi(\mu)(\xi_n - \mu)$$

$$\text{Var}(\dots) \approx \text{Var}(\nabla\varphi(\mu)(\xi_n - \mu)) = \text{Var}(\nabla\varphi(\mu)\xi_n) = \nabla\varphi \text{Var} \xi_n (\nabla\varphi)^T$$

Вернемся к следующему равенству:

$$\varphi(\xi_n) - \varphi(\mu) \approx \nabla\varphi(\mu)(\xi_n - \mu)$$

$$\sqrt{n}(\varphi(\xi_n) - \varphi(\mu)) \approx \sqrt{n}\nabla\varphi(\mu)(\xi_n - \mu)$$

$$\sqrt{n}(\xi_n - \mu) \rightarrow N(0, \Sigma) \quad (\text{Как мы и говорили выше})$$

А мы домножаем вектор на градиент, поэтому:

$$\sqrt{n}\nabla\varphi(\mu)(\xi_n - \mu) \rightarrow N(0, \nabla\varphi(\mu)\Sigma(\nabla\varphi(\mu))^T)$$

Это называется дельта метод.

Теорема

Пусть $\xi_n = (\bar{X}, \dots, \bar{X}^k)$

Многомерная версия ЦПТ

$$\sqrt{n}(\xi_n - \alpha) \xrightarrow{d} N(0, \Sigma)$$

$$\alpha = (\alpha_1, \dots, \alpha_k)$$

$$\Sigma = \text{Var}(X_1, \dots, X_1^k)$$

$$\varphi: \mathbb{R}^k \rightarrow \mathbb{R}$$

$$\varphi \in C^1(\mathbb{R}^k)$$

$$\sigma^2 = \nabla\varphi(\alpha)(\nabla\varphi(\alpha))^T > 0$$

Тогда (продолжение теоремы):

$$\sqrt{n} \frac{\varphi(\xi_n) - \varphi(\alpha)}{\sigma} \xrightarrow{d} N(0, 1)$$

Кроме того:

$$\sigma = \sigma(\alpha) \in C^1(\mathbb{R}^k) \Rightarrow \sqrt{n} \frac{\varphi(\xi_n) - \varphi(\alpha)}{\sigma(\xi_n)} \xrightarrow{d} N(0, 1)$$

$$\text{Коэффициент асимметрии: } \frac{E(X - EX)^3}{\sigma^3} = \gamma$$

$$\frac{\widehat{\beta_3}}{\widehat{S_*^3}} = \widehat{\gamma}$$

$$\text{Коэффициент эксцесса: } \frac{E(X - EX)^4}{\sigma^4} - 3$$

$$\frac{\widehat{\beta_4}}{\widehat{S_*^4}} - 3$$

Пусть у нас есть две выборки:

$$X_1, \dots, X_n$$

$$Y_1, \dots, Y_n$$

$$\text{Cov}(X, Y) = EXY - EX \cdot EY = E(X - EX)(Y - EY)$$

$$S_{*XY} = \frac{1}{n} \sum_j (X_j - \bar{X})(Y_j - \bar{Y}) = \frac{1}{n} \sum_j X_j Y_j - \bar{X}\bar{Y}$$

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var} X \cdot \text{Var} Y}}$$

$$\rho = \frac{S_{*XY}}{S_{*X} \cdot S_{*Y}} - \text{Выборочный коэффициент корреляции (коэффициент корреляции Пирсона)}$$

5 Порядковые статистики

Определение. Вариационный ряд

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ - вариационный ряд

Определение. Порядковая статистика

$X_{(k)}$ - k -я порядковая статистика.

Квантиль порядка α

$$P(X \geq q_\alpha) \geq 1 - \alpha$$

$$P(X \leq \alpha) \geq \alpha$$

Это общее определение

Если F строго возрастает:

$$F(q_\alpha) = \alpha \Leftrightarrow q_\alpha = F^{-1}(\alpha)$$

$$F^{-1}(\alpha) : \sup\{x : F(x) \leq \alpha\}, \inf\{x : F(x) \geq \alpha\}$$

Определение. Выборочный квантиль порядка α

$$\alpha \in (0, 1)$$

$$\exists 0 \leq k \leq n-1 : \frac{k}{n} \leq \alpha < \frac{k+1}{n}$$

$X_{(k+1)}$ - выборочный квантиль порядка α

$\alpha = 0$ - нулевой квантиль

$$F^{-1} = \sum \{x \in \mathbb{R} F_n(x) \leq \alpha\}$$

$\alpha = \frac{1}{4}$ - первый квантиль (нижний квантиль)

$\alpha = \frac{1}{2}$ - второй квантиль (выборочная медиана)

$\alpha = \frac{3}{4}$ - третий квантиль (верхний квантиль)

$\alpha = 1$ - $\max(X)$ (четвертый квантиль)

$$n = 2m \Rightarrow \text{med}(X) = \frac{X_{(m)} + X_{(m+1)}}{2}$$

$$n = 2m + 1 \Rightarrow \text{med}(X) = X_{(m+1)}$$

$IQR = \Delta$ между верхним и нижним квантилем.

$$P(X_{(k)} \leq t) = P(\mu_n(t) \geq k) = \sum_{j=k}^n C_n^j F^j(t) (1 - F(t))^{n-j}$$

$$B(z, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z t^{a-1} (1-t)^{b-1} dt = B(F(t), k, n-k+1)$$

$$0 \leq z \leq 1$$

Пусть $p()$ - теоретическая плотность, то есть $p = F'$

$$(P(X_{(k)} \leq t))'_t = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \cdot F^{k-1}(t)(1-F(t))^{n-k} \cdot p(t) - \text{плотность}$$

k -й порядковой статистики

Рассуждая аналогично можно получить плотность для двухмерного случая:

$$g(x_1, x_2) = \frac{n!}{(k-1)!(r-k-1)!(n-k)!} \cdot F^{k-1}(x_1)(F(x_2) - F(x_1))^{r-k-1}(1 -$$

$F(x_2))^{n-r} p(x_1)p(x_2)$ - плотность для $(X_{(k)}, X_{(r)})$, $l < r$

$g(x_1, \dots, x_n) = n!p(x_1) \cdot \dots \cdot p(x_n)$ - плотность для $(X_{(1)}, \dots, X_{(n)})$

Средний член вариационного ряда: $\frac{K(n)}{n} \rightarrow const \in (0, 1)$

Крайний член вариационного ряда:

$X_{(r)}$, r - огр.

$X_{(n+1-s)}$, s - огр.

Теорема об асимптотике среднего члена вариационного ряда

$0 < \alpha < 1$ - теоретическая плотность.

q_α - теоретический квантиль порядка α

$p \in C^1$ (окр-сть q_α)

$p(q_\alpha) > 0$

Тогда:

$$\sqrt{n} \cdot f(q_\alpha) \frac{X_{(\lfloor n\alpha \rfloor)} - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} N(0, 1)$$

Идея доказательства

Пусть $\lfloor n\alpha \rfloor = k$

Мы умеем писать плотность для $X_{(k)}$

Затем у нас идет преобразование:

$$g(x) = \sqrt{np}(q_\alpha) \frac{x - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \rightsquigarrow p_{g(X_{(k)})}(t) = p_{X_{(k)}}(g^{-1}(t)) |g^{-1}(t)'_t| \text{ (теорема из}$$

прошлого семестра)

Там вылезут факториалы, от них мы умеем избавляться по Стирлингу

Затем надо будет воспользоваться непрерывной дифференцируемостью:

$p \in C^1$ (окр-сть q_α)

$p(q_\alpha) > 0$

Тогда в пределе наша новая плотность будет стремиться к плотности нормального стандартного закона.

Пример. Распределение Коши.

$$Cauchy(\mu, 1) \quad \alpha = \frac{1}{2}$$

$$2f(\mu)\sqrt{n}(X_{(\lfloor \frac{n}{2} \rfloor)}) = \frac{1}{\pi} \cdot \frac{1}{1 + (\mu - \mu)^2} \cdot 2\sqrt{n}(X_{(\lfloor \frac{n}{2} \rfloor)} - \mu)$$

Теорема об асимптотике крайних членов вариационного ряда

$r, s, F, x, p()$ - плотность

Тогда:

$$nF(X_{(r)}) \xrightarrow{d} \Gamma(r, 1)$$

$$nF(X_{(n+1-s)}) \xrightarrow{d} \Gamma(s, 1)$$

И оба распределения независимы.

Идея доказательства

У нас есть совместная плотность и какое-то преобразование, тогда мы можем написать плотность после преобразования

Затем берем предел и мы получим плотность равная произведению двух этих двух законов.

6 Постановка задачи точечного оценивания параметров

$X_1, \dots, X_n \sim F_\theta \in \Theta \subset \mathbb{R}^d$

θ - некий фиксированный неизвестный вектор.

Наша цель оценить θ в виде $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$

Замечание:

1. $\hat{\theta}(X_1, \dots, X_n)$ - статистика (измеримая функция от выборки)
2. Байесовская постановка: θ - случайная величина из известного априорного распределения

Определение. Состоятельность

$\hat{\theta}$ - состоятельная оценка $\theta \Leftrightarrow \hat{\theta} \xrightarrow{P} \theta$

Определение. Несмещенность

$b.as(\hat{\theta}) \stackrel{def}{=} E\hat{\theta} - \theta$ - смещение

$b.as(\hat{\theta}) = 0 \Leftrightarrow$ несмещенная

Определение. Асимптотическая нормальность

$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, \Sigma_\theta)$

Пример

1. $X_1, \dots, X_n \sim \text{Bern}(p)$
 $\hat{\theta} = \bar{X}$
ЗБЧ $\bar{X} \rightarrow p$
 $E\hat{\theta} = E\bar{X} = p$ (несм)
ЦПТ $\sqrt{n} \frac{\bar{X} - p}{\sqrt{p(1-p)}} \rightarrow N(0, 1)$
2. $N(\mu, \sigma^2)$, μ - известна
 $\theta = S_*^2$
 $S_*^2 \rightarrow \sigma^2$
 $ES_*^2 = \frac{n-1}{n} \sigma^2 \neq \sigma^2$ (несмещенность)
 $\sqrt{n} \frac{S_*^2 - \sigma^2}{\sqrt{\beta_4 - \sigma^4}} \xrightarrow{d} N(0, 1)$

Определение. Эффективность (оптимальность)

$\hat{\theta}_1$ эффективнее $\hat{\theta}_2 \Leftrightarrow MSE\hat{\theta}_1 < MSE\hat{\theta}_2$

$MSE\hat{\theta} = E \|\hat{\theta} - \theta\|^2 = E(\hat{\theta} - \theta)^T (\hat{\theta} - \theta)$

Утверждение

$MSE\hat{\theta} = \text{tr}(\text{Var} \hat{\theta}) + \|b.as\hat{\theta}\|^2$

Доказательство

$MSE = E(\hat{\theta} - \theta)^T (\hat{\theta} - \theta) = E(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)^T (\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta) = E(\hat{\theta} - E\hat{\theta})^T (\hat{\theta} - E\hat{\theta}) + \sum \text{Var} \hat{\theta}_i + \|b.as\hat{\theta}\|^2$

1. Ассимптотическая нормальность \Rightarrow состоятельность

$$\hat{\theta} - \theta = \frac{1}{\sqrt{n}}(\hat{\theta} - \theta) \xrightarrow{P} 0$$

2. Ассимптотическая нормальность $\Rightarrow b.as\hat{\theta} \rightarrow 0$

Пусть $d = 1$

$$P(|\hat{\theta} - E\hat{\theta}| > \varepsilon) = P\left(\frac{\sqrt{n}|\hat{\theta} - E\hat{\theta}|}{\sigma} > \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - P\left(\dots < \frac{\varepsilon\sqrt{n}}{\sigma}\right) \approx$$

$$1 - (2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1) = 2(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)) \rightarrow 0$$

3. Состоятельность $\Rightarrow b.as\hat{\theta} \rightarrow 0$

Следует из усиленного закона больших чисел

$\bar{X} \xrightarrow{a.s.} \mu \Rightarrow E\bar{X} \rightarrow \mu$ (По теореме Лебега о мажорируемой сходимости)

4. Пусть $d = 1$, $b.as\hat{\theta} \rightarrow 0$, $\text{Var } \hat{\theta} \rightarrow 0 \Rightarrow \hat{\theta}$ - сост.

7 Метод моментов

Рассмотрим g_1, \dots, g_d

$$\exists E g_1(X_1) = m_1(\theta_1, \dots, \theta_d)$$

$$\exists E g_2(X_2) = m_2(\theta_1, \dots, \theta_d)$$

...

$$\exists E g_d(X_d) = m_d(\theta_1, \dots, \theta_d)$$

$$\begin{cases} \overline{g_1(X)} = m_1(\hat{\theta}_1, \dots, \hat{\theta}_d) \\ \dots \\ \overline{g_d(X)} = m_d(\hat{\theta}_1, \dots, \hat{\theta}_d) \end{cases}$$

Пусть $\exists!$ решение:

$$\begin{cases} \hat{\theta}_1 = \alpha_1(\overline{g_1(X)}, \dots, \overline{g_d(X)}) \\ \dots \\ \hat{\theta}_d = \alpha_d(\overline{g_1(X)}, \dots, \overline{g_d(X)}) \end{cases}$$

Тогда это будет оценка методов моментов.

По умолчанию $g_j(x) = x^j$

Пример

$$U[\theta_1, \theta_2], \theta_1 < \theta_2$$

$$g_1(x) = x, g_2(x) = x^2$$

$$E g_1(X_1) = E X_1 = \frac{\theta_1 + \theta_2}{2}$$

$$E g_2(X_1) = E X_1^2 = \frac{(\theta_2 - \theta_1)^2}{12} + \frac{(\theta_1 + \theta_2)^2}{4} = \frac{\theta_2^2 - 2\theta_1\theta_2 + \theta_1^2}{12} + \frac{\theta_1^2 + 2\theta_1\theta_2 + \theta_2^2}{4} =$$

$$\frac{\theta_1^2 + \theta_1\theta_2 + \theta_2^2}{3}$$

Далее составим уравнения:

$$\begin{cases} \bar{X} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2} \\ \bar{X^2} = \frac{\hat{\theta}_1^2 + \widehat{-\theta_1\theta_2} + \hat{\theta}_2^2}{3} \end{cases} \Leftrightarrow \begin{cases} \hat{\theta}_2 = 2\bar{X} - \hat{\theta}_1 \\ 3\bar{X^2} = \hat{\theta}_1^2 + \widehat{-\theta_1\theta_2} + \hat{\theta}_2^2 \end{cases}$$

$$3\bar{X^2} = \hat{\theta}_1^2 + \hat{\theta}_1(2\bar{X} - \hat{\theta}_1) + (2\bar{X} - \hat{\theta}_1)^2$$

$$\hat{\theta}_1^2 - 2\bar{X}\hat{\theta}_1 + 4(\bar{X}^2 - 3\bar{X^2}) = 0$$

Считаем дискриминант деленный на четыре:

$$\frac{D}{4} = (\bar{X})^2 - 4(\bar{X}^2 - 3\bar{X^2}) = 3(\bar{X^2} - (\bar{X})^2) = 3S_*^2$$

У нас будет два случая:

1. $\begin{cases} \hat{\theta}_1 = \bar{X} + \sqrt{3}S_* \\ \hat{\theta}_2 = \bar{X} - \sqrt{3}S_* \end{cases}$
2. $\begin{cases} \hat{\theta}_1 = \bar{X} - \sqrt{3}S_* \\ \hat{\theta}_2 = \bar{X} + \sqrt{3}S_* \end{cases}$

Первый не возможен, потому что $\theta_1 < \theta_2$

1. если $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$ - состоятельная оценка
2. если $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$ - асимптотически нормальные и g_1, \dots, g_d - гладкие, то каждая из оценок асимптотически нормальные.

8 Метод максимального правдоподобия

probability must function: $p(x, \theta) = p(x|\theta)$

probability identity function: $p(x, \theta) = p(x|\theta)$

Будем называть оба случая плотностью.

Пусть у нас есть выборка $X_1, \dots, X_n \sim p(x|\theta)$

$L(x|\theta) = \prod p(x_i|\theta)$ - функция правдоподобия

$\theta^* = \underset{\hat{\theta}}{argmax}(L(x, \theta))$ - оценка максимума правдоподобия

$\theta \in \Theta$ - открыто

$\theta_1 \neq \theta_2 \Rightarrow L(x, \theta_1) \neq L(x, \theta_2)$

Доказательство

1. Посмотреть и подумать
2. Рассмотреть $\ln L(x|\theta)$; $\frac{\partial \ln L(x, \theta)}{\partial \theta}$
3. $\frac{\partial \ln L(x, \theta)}{\partial \theta} = 0$
4. Проверить достаточные условия максимума

Пример

1. $N(\theta_1, \theta_2)$

$$L(x, \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x_i - \theta_1)^2}{2\theta_2}\right)$$

$$\ln L(x, \theta) = \sum_{i=1}^n \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2}\theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2}\right]$$

$$\frac{\partial \ln L(x, \theta_1)}{\partial \theta_1} = \sum_{i=1}^n \frac{2(x_i - \theta_1)}{2\theta_2} = \sum_{i=1}^n \frac{x_i - \theta_1}{\theta_2}$$

$$\frac{\partial \ln L(x, \theta)}{\partial \theta_2} = \sum_{i=1}^n \left[-\frac{1}{2\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2}\right]$$

$$\sum_{i=1}^n \frac{(x_i - \hat{\theta}_1)}{\hat{\theta}_2} = 0 \Rightarrow \hat{\theta}_1 = \bar{X}$$

$$\sum_{i=1}^n \left[-\frac{1}{2\hat{\theta}_2} + \frac{(x_i - \hat{\theta}_1)^2}{2\hat{\theta}_2^2}\right] = 0 \Rightarrow \hat{\theta}_2 = S_*^2$$

2. $U[0, \theta] : L(X, \theta) = \frac{1}{\theta^n} \mathbb{1}(X_i \in [0, \theta], \forall 1 \leq i \leq n)$

$$X_1, \dots, X_n : p(X_i, \theta)$$

$$L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n p(x_i, \theta)$$

$$L(X, \theta) = \theta^{\sum X_i} (1 - \theta)^{n - \sum X_i}$$

Вернемся к нашему примеру:

Если $\hat{\theta} < \max(X)$, то $L(x, \hat{\theta}) = 0$

Чем меньше θ , тем больше $\frac{1}{\theta}$

$$\Rightarrow \hat{\theta} = \max(X)$$

$$Poly(1, p) : p = (p_1, \dots, p_m)$$

Рассмотрим частоты:

ν_1 - кол-во наблюдений типа 1

...

ν_m - кол-во наблюдений типа m

Суммируем и смотрим на функцию правдоподобия

$$L(X, p) = p_1 \dots p_m$$

$$\ln L(X, p) = \sum_{j=1}^{m-1} \nu_j \ln p_j + \nu_m \ln (1 - p_1 - \dots - p_{m-1})$$

$$\frac{\partial \ln L \dots}{\partial p_j} = \frac{\nu_j}{p_j} - \frac{\nu_m}{1 - p_1 - \dots - p_{m-1}} = 0$$

$$\sum \text{уравнения: } \nu_j (1 - \hat{p}_1 - \dots - \widehat{p_{m-1}}) = \hat{p}_j \cdot \nu_m$$

$$\widehat{p_m} (n - \nu_m) = \nu_m (1 - \widehat{p_m})$$

$$\widehat{p_m} n - \widehat{p_m} \nu_m = \nu_m$$

$$\widehat{p}_m = \frac{\nu_m}{n}$$

$$\widehat{p}_j = \frac{\nu_j \widehat{p}_m}{\nu_m} = \frac{\nu_j}{n}$$

9 Информация Фишера

$$d = 1 : L(X, \theta) = \prod p(X_j, \theta)$$

$$\ln L(X, \theta) = \sum \ln p(x_j, \theta)$$

$$V(X, \theta) = \frac{\partial \ln L \dots}{\partial \theta} = \sum \frac{\partial \ln p \dots}{\partial \theta} - \text{вклад выборки}$$

$\theta \in \Theta$ - открыто

$$\theta_1 \neq \theta_2 \Rightarrow p(X, \theta_1) \neq p(X, \theta_2)$$

Регулярность:

$$1. \frac{\partial}{\partial \theta} \int_X T(X) L(X_i, \theta) dX = \int \frac{\partial}{\partial \theta} L(X, \theta) \cdot T(X) dX$$

Необходимое условие $\sup p_x$ не зависит от θ

$$U[0, \theta] \int_0^\theta \frac{1}{\theta} dt = 1$$

$$(\int_0^\theta \frac{1}{\theta} dt)'_\theta = (\frac{1}{\theta} \int_0^\theta dt)'_\theta = -\frac{1}{\theta^2} \int_0^\theta dt + \frac{1}{\theta} = 0 \neq \int_0^\theta (\frac{1}{\theta})'_\theta dt$$

$$2. EV^2(X, \theta) < \infty$$

$$\int_X L(X, \theta) dX = 1 \xrightarrow{\frac{\partial}{\partial \theta}} \int_X \frac{\partial L(\cdot)}{\partial \theta} dX = \int_X \frac{\frac{\partial L(\dots)}{\partial \theta}}{L(\dots)} \cdot L(\dots) dX = \int_X V(X, \theta) L(X, \theta) dX =$$

$$EV(X, \theta) = 0$$

$I(\theta) = \text{Var}(V(X_i, \theta)) = E(V^2(X_i, \theta))$ - информация Фишера для всей выборки

$$V(X, \theta) = \sum_j \frac{\partial \ln p(x, \theta)}{\partial \theta} \Rightarrow \text{Var}(V(X, \theta)) = n \cdot \text{Var} \frac{\partial \ln p(x, \theta)}{\partial \theta}$$

$i(\theta)$ - информация Фишера для 1 наблюдения

$$i(\theta) = E\left(\frac{\partial \ln p(x_j, \theta)}{\partial \theta}\right)^2$$

$$\frac{\partial}{\partial \theta} \int_{\mathbb{R}} \frac{\partial \ln P(x, \theta)}{\partial \theta} \cdot p(x, \theta) dx = \int_{\mathbb{R}} \frac{\partial^2 \ln p(x, \theta)}{\partial \theta} dx + \int_{\mathbb{R}} \frac{\partial \ln \dots}{\partial \theta} \frac{\partial p \dots}{\partial \theta} dx = E \frac{\partial^2 \ln p(x, \theta)}{\partial \theta^2} +$$

$$E\left(\frac{\partial \ln o(x, \theta)}{\partial \theta}\right)^2 = 0$$

$$i(\theta) = E\left(\frac{\partial \ln p(x_j, \theta)}{\partial \theta}\right)^2 = -E \frac{\partial^2 \ln p(x, \theta)}{\partial \theta^2}$$

Произвольное d :

$$i(\theta) = -(E \frac{\partial^2 \ln p(X, \theta)}{\partial \theta_i \partial \theta_j})_{1 \leq i, j \leq d}$$

$$I(\theta) = ni(\theta)$$

$$N(\theta_1, \theta_2)$$

$$p(x, \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} \cdot \exp\left(-\frac{(x - \theta_1)^2}{2\theta_2}\right)$$

$$\ln p(x, \theta_1, \theta_2) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \theta_2 - \frac{(x - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial \ln p(x, \theta_1, \theta_2)}{\partial \theta_1} = \frac{x - \theta_1}{\theta_2}$$

$$\frac{\partial \dots}{\partial \theta_2} = -\frac{1}{2\theta_2} + \frac{(x - \theta_1)^2}{\theta_2^2} \text{ TODO:}$$

Теорема. Неравенство Рао-Крамера

Модель регулярная, $d = 1$

$\tau(\theta)$ - оцениваемая функция

$\tau \in C^1$ (как правило $\tau(\theta) = \theta$)

$$\widehat{\tau(\theta)} = \tau(\theta) \Rightarrow \text{Var } \widehat{\tau(\theta)} \geq \frac{[\tau'(\theta)]^2}{ni(\theta)}$$

$$\tau'(\theta) = \int \widehat{\tau(\theta)} \frac{\partial L(X, \theta)}{\partial \theta} dX = \int \widehat{\tau(\theta)} V(X, \theta) dX - EV(X, \theta) \cdot E\widehat{\tau(\theta)} = \text{Cov}(V(X, \theta), \widehat{\tau(\theta)})$$

$$\text{Cov}^2(V(X, \theta), \widehat{\tau(\theta)}) \leq \text{Var}(V(X, \theta)) \cdot \text{Var}(\widehat{\tau(\theta)})$$

Замечания

$$1. \quad E\widehat{\tau(\theta)} - \tau(\theta) = \text{bias}(\theta) \neq 0$$

$$E\widehat{\tau(\theta)} = \tau(\theta) + \text{bias}(\theta)$$

$$\text{Var } \widehat{\tau(\theta)} \geq \frac{[\tau'(\theta) + \text{bias}'(\theta)]^2}{ni(\theta)}$$

2.

Многомерный случай

$$\tau(\theta) : \mathbb{R}^d \rightarrow \mathbb{R}$$

$$\tau \in C^1$$

$$E\widehat{\tau\theta} = \tau\theta \Rightarrow \text{Var } \widehat{\tau(\theta)} \geq \frac{\nabla \tau(\theta) i^{-1}(\theta) \nabla^T \tau(\theta)}{n}$$

Пример

$$N(\theta_1, \theta_2)$$

$$\tau(\theta_1, \theta_2) = \theta_1$$

$$E\widehat{\theta_1} = \theta_1 \Rightarrow \text{Var } \widehat{\theta_1} \geq \frac{(1, 0) \begin{pmatrix} \theta_2 & 0 \\ 0 & 2\theta_2^2 \end{pmatrix} \cdot (1, 0)^T}{n} = \frac{\theta_2}{n}$$

Свойства оценки

1. Если \exists несмещенная оптимальная оценка в регулярном случае то она совпадает с оценкой максимального правдоподобия (ОМП)

$$\tau(\theta) = \theta$$

$$V(X, \theta) = \frac{1}{a(\theta)}(\widehat{\theta} - \theta)$$

10 Состоятельность ОМП

Пусть θ_0 - реальный параметр $\Rightarrow p_{\theta_0}(L(X, \theta_0) > L(X, \theta)) \rightarrow 1$

$$\frac{L(X, \theta)}{L(X, \theta_0)} < 1$$

$$\frac{1}{n} \sum \ln \frac{p(X_j, \theta)}{p(X_j, \theta_0)} < 0$$

$$\text{По ЗБЧ} \Rightarrow E_{\theta_0} \ln \frac{p(x_j, \theta)}{p(x_j, \theta_0)} \leq E_{\theta_0} \left[\frac{p(X_j, \theta)}{p(X_j, \theta_0)} - 1 \right] = \int_X p(X, \theta) dX - \int p(X, \theta_0) dX = 0$$

Давайте введем события:

$$S_n = \{X : \ln L(X, \theta_0) > \ln L(X, \theta_0 - a)\} \cap \{X : \ln L(X, \theta_0) > \ln L(X, \theta_0 + a)\}$$

$$P_{\theta_0}(S_n) \rightarrow 1$$

$$A_n = \{X : |\hat{\theta} - \theta_0| < a\}$$

$$B_n = \{X : \frac{\partial \ln L(X, \theta)}{\partial \theta} \big|_{\theta=\hat{\theta}} = 0\}$$

$$S_n \subset A_n B_n \subset A_n \Rightarrow P(A_n) \rightarrow 1$$

Давайте поговорим про свойства метода максимального правдоподобия

1. Принцип инвариантности

$$\theta \in \Theta \xrightarrow{\text{bijection}} \gamma \in \Gamma$$

$$\theta = \varphi^{-1}(\gamma) \Leftrightarrow \gamma = \varphi(\theta)$$

$$\sup_{\theta} L(X, \varphi(\gamma)) = \sup_{\gamma} L(x, \gamma)$$

$$\gamma^* = \varphi(\theta^*)$$

Пример

Пусть у нас есть $Exp(\lambda)$ и есть две параметризации

- $\lambda e^{-\lambda x} \rightarrow \frac{1}{\bar{X}}$
- $\frac{1}{\lambda} \exp(-\frac{x}{\lambda}) \rightarrow \bar{X}$

Теорема Асимптотическая нормальность ОМП

Пусть наша модель регулярная, так же пусть:

$$\left| \frac{\partial^3 \ln f(x, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq M$$

θ_* - ОПМ для θ

Уравнение $\nabla \ln L(X, \theta) = 0$ имеет единственное решение. Тогда:

1. $\sqrt{n}(\theta_* - \theta) \rightarrow N(0, i^{-1}(\theta))$
2. $\tau(\theta)$ - оцениваемая функция от θ
 $\tau \in C^1$
 $\sqrt{n}(\tau(\theta_*) - \tau(\theta)) \rightarrow N(0, \sigma^2)$
 $\sigma^2 = \nabla \tau(\theta) i^{-1}(\theta) \nabla^T \tau(\theta)$

3. σ^2 - непрерывная функция от $\theta \Rightarrow \sqrt{\frac{\tau(\theta_*) - \tau(\theta)}{\sigma(\theta_*)}} \rightarrow N(0, 1)$

В прошлый раз мы ввели функцию

$$V(X, \theta) = \frac{\partial \ln L(X, \theta)}{\partial \theta}$$

θ_0 - реальный параметр

Давайте напишем ряд Тейлора

$$V(X, \theta) = V(X, \theta_0) + V'_\theta(X, \theta)(\theta - \theta_0) + V''_\theta(X, \tilde{\theta}) \frac{(\theta - \theta_0)^2}{2}, \tilde{\theta} \text{ между } \theta_0 \text{ и } \theta$$

Выполним подстановку $\theta = \theta_*$

$$0 = V(X, \theta_0) + V'_\theta(X, \theta_0)(\theta_* - \theta_0) + V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2}$$

$$V'_\theta(X, \theta_0)(\theta_* - \theta_0) = -V(X, \theta_0) - V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2}$$

$$\sqrt{n}V'_\theta(X, \theta_0)(\theta_* - \theta_0) = -\sqrt{n}V(X, \theta_0) - \sqrt{n}V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2}$$

$$A_n := -\sqrt{n}V(X, \theta_0)$$

По ЦПТ:

$$A_n \rightarrow N(0, i(\theta_0))$$

$$\sqrt{n}V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2} = n^{\frac{3}{2}} - \frac{V''_\theta(X, \tilde{\theta}) (\theta_* - \theta_0)^2}{n}$$

$$\frac{V''_\theta(X, \tilde{\theta})}{n} - \text{огр по ЗБЧ}$$

$$\sqrt{n}V''_\theta(X, \tilde{\theta}) \frac{(\theta_* - \theta_0)^2}{2} \rightarrow N(0, i(\theta_0))$$

$$V'(X, \theta_0) = n \frac{V'(X, \theta_0)}{n} \xrightarrow{\text{ЗБЧ}} -i(\theta)$$

$$\text{Var } \hat{\theta} \geq \frac{[\tau'(\theta)]^2}{ni(\theta)}$$

Рассмотрим показатель:

$$\frac{[\tau'(\theta)]^2}{ni(\theta) \cdot \widehat{\text{Var } \tau(\theta)}} - \text{Эффективность}$$

Асимптотическая Эффективность: Пусть $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, \frac{\sigma^2}{n})$

$$\frac{1}{i(\theta)\sigma^2} - \text{показатель состоятельной эффективности}$$

11 Экспоненциальное семейство распределений

Пусть наше распределение относится к экспоненциальному семейству распределений если:

$$p(x, \theta) = \exp(A(\theta)B(x) + C(\theta) + D(x))$$

К таким распределениям относятся: $N(), \Gamma(), Pois(), Bin, NB$

$$\ln p(x, \theta) = A(\theta)B(x) + C(\theta) + D(x)$$

$$\begin{aligned}\frac{\partial \ln p(x, \theta)}{\partial \theta} &= A'(\theta)B(x) + C'(\theta) \\ V(X, \theta) &= A'(\theta) \sum B(X_i) + nC'(\theta) \\ V(X, \theta) &= n(A'(\theta)\overline{B(X)} + C''(\theta)) \\ \frac{V(X, \theta)}{n} - C'(\theta) &= A'(\theta)\overline{B(X)} \\ \overline{B(X)} &= \frac{V(X, \theta)}{nA'(\theta)} - \frac{C'(\theta)}{A'(\theta)} \\ \overline{B(X)} &- \text{оптимальная оценка для } \left(-\frac{C'(\theta)}{A'(\theta)}\right)\end{aligned}$$

12 Байесовская постановка

$$\begin{aligned}X_1, \dots, X_n &\sim F_\theta \\ \theta &\sim \pi(\theta) - \text{prior} \\ l(\hat{\theta}, \theta) &- \text{функция потерь} \\ l(\hat{\theta}, \theta) &= (\hat{\theta} - \theta)^2 \text{ (default)} \\ R(\hat{\theta}, \theta) &= El(\hat{\theta}, \theta) - \text{риск} \\ r(\hat{\theta}) &= E_{\pi(\theta)} R(\hat{\theta}, \theta) - \text{байесовский риск} \\ \hat{\theta}_B &= \underset{\theta}{\operatorname{argmin}} r(\hat{\theta}) \\ r(\hat{\theta}) &= El(\hat{\theta}, \theta) \\ \text{Давайте вспомним теорему Байеса:} \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ P(\theta|X) &= \frac{L(X|\theta)\pi(\theta)}{\int L(X|\theta)\pi(\theta)d\theta} \\ \text{posterior} &= \text{likelihood} \times \text{prior} \\ \hat{\theta}_B &= \underset{\hat{\theta}}{\operatorname{argmin}} E[l(\hat{\theta})|X] \\ r(\theta_*) &\leq r(\hat{\theta}) \\ \text{TODO: как будет не лень я допишу пример 1:05 лекция 6} \\ l(\hat{\theta}, \theta) &- \text{loss function} \\ R(\hat{\theta}, \theta) &= E_\theta l(\hat{\theta}, \theta) \\ r(\hat{\theta}) &= E_{\pi(\theta)} R(\hat{\theta}, \theta) \\ \hat{\theta}_B &= \underset{\hat{\theta}}{\operatorname{argmin}} E(l(\hat{\theta}, \theta)|X) \\ l(\hat{\theta}, \theta) &= (\hat{\theta} - \theta)^2 \\ \hat{\theta}_B &= E(\theta|X)\end{aligned}$$

13 Минимаксные оценки

$$m(\hat{\theta}) = \sup_{\theta} R(\hat{\theta}, \theta)$$

$$\hat{\theta}_{WC} = \operatorname{argmin}_{\theta} m(\hat{\theta}) - \text{минимаксная оценка}$$

$$r(\hat{\theta}) \leq m(\hat{\theta})$$

Утверждение

$$\exists \pi(\theta) - \text{prior} : R(\hat{\theta}_B, \theta) = \text{const} \Rightarrow \hat{\theta}_{WC} = \hat{\theta}_B$$

Рассмотрим пример:

$$\text{Bern}(p) : \text{prior: } B(a, b)$$

$$\hat{p}_B = \frac{a + X}{a + b + n}, X = \sum_{k=1}^n X_k$$

$$R(\hat{p}_B, p) = MSE \hat{p}_B = E(\hat{p}_B - p)^2 = \text{Var } \hat{p}_B + bias^2 \hat{p}_B$$

$$bias \hat{p}_B = E \frac{a + X}{a + b + n} - p = \frac{a + np}{a + b + n} - p = \frac{a - ap - bp}{a + b + n}$$

$$\text{Var } \hat{p}_B = \frac{1}{(a + b + n)^2} \text{Var } X = \frac{npq}{(a + b + n)^2}$$

$$R(\hat{p}_B, p) = \frac{(a - ap - bp)^2 + npq}{(a + b + n)^2} = \frac{(a - p(a + b))^2 + bp(1 - p)}{(a + b + n)^2} = \frac{p^2((a + b)^2 - n) + p(-2a(a + b) + n) + a^2}{(a + b + n)^2}$$

$$\begin{cases} (a + b)^2 = n \\ 2a(a + b) = n \end{cases} = \begin{cases} a + b = \sqrt{n} \\ a = \frac{\sqrt{n}}{2} \end{cases} \Rightarrow b = \frac{\sqrt{n}}{2}$$

- Достаточность и полные статистики
- Робастность (устойчивость относительно выбросов, устойчивость относительно изменения параметров распределения)

14 Интервальное оценивание

Определение. Доверительный интервал

$$X_1, \dots, X_n \sim F_{\theta}, \theta \in \Theta \subset \mathbb{R}$$

$1 - \alpha = \gamma \in (0, 1)$ - уровень доверия

default: 0.9, 0.95, 0.99

$(T_l(X), T_r(X))$ - доверительный интервал уровня $\gamma = 1 - \alpha$ если $p(\theta \in$

$(T_l(X), T_r(X)) \geq \gamma)$

Пусть $T(X, \theta) \sim G$ - не зависит от θ

Рассмотрим $p(q_1 < T(X, \theta) < q_2) = 1 - \alpha$

$$q_1 = q_{\frac{\alpha}{2}}$$

$$q_2 = q_{1 - \frac{\alpha}{2}}$$

Универсальный рецепт (нет)

а) $F_{\theta}(X_k)$

$$P(F_{\theta}(X_k) \leq t) = P(X_k \leq F_{\theta}^{-1}(t)) = F_{\theta}(F_{\theta}^{-1}(t)) = t$$

б) $-\ln F_{\theta}(X_k) \sim \text{Exp}(1)$

$$p(-\ln U \leq t) = p(U \geq e^{-t}) = 1 - e^{-t}$$

$$v) - \sum \ln F_{\theta}(X_k) \sim \Gamma(n, 1)$$

Доверительные интервалы нормального закона. Теорема Фишера

Лемма о независимости линейной и квадратичной статистик

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$$T = AX, X = (X_1, \dots, X_n)^T, A \in M_{m \times n}(\mathbb{R})$$

$$Q = X^T B X, B \in M_n(\mathbb{R}), B = B^T$$

Тогда T, Q - независимы

$$AB = 0$$

Доказательство

$$\Lambda = U^T B U$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m, 0, 0)$$

λ_k - собственное число не 0

$U = (u_1, \dots, u_n)$ - собственные векторы ортонормированного базиса $\Leftrightarrow B =$

$$U \Lambda U^T = \sum_{j=1}^m \lambda_j u_j u_j^T \Rightarrow Q = \sum_{j=1}^M \lambda_j (X^T U_j) (U_j^T X) = \sum_j \lambda_j (U_j^T X)^2$$

$$A \left(\sum_{j=1}^m \lambda_j u_j u_j^T \right) = 0$$

$$\sum_{j=1}^m \lambda_j A U_j u_j^T = 0$$

Зафиксируем $1 \leq k \leq m$ домножим справа на u_k

$$A u_k = 0 \Rightarrow \forall i, k A[i, *] u_k = 0$$

Нам надо доказать, что $\forall i, k A[i, *] X$ и $u_k^T X$ - нез

$$\text{Cov}(A[i, *] X, u_k^T X) = \text{Cov}(A[0, *] X, X^T u_k) = A[i, *] \text{Var } X u_k = \sigma^2 A[i, *] u_k = 0$$

Лемма о независимости двух квадратичных статистик

$$Q_1 = X^T B_1 X$$

$$Q_2 = X^T B_2 X$$

Тогда Q_1, Q_2 - нез

$$B_1 B_2 = B_2 B_1 = 0$$

Определение X_n - квадратичная

$$X_1, \dots, X_n \sim N(0, 1)$$

$$\sum_{k=1}^n X_k^2 \sim \chi^2(n) \text{ (распределение)}$$

хи-кквadrat с n степенями свободы

Лемма о распределении квадратичной статистики

$$X_1, \dots, X_n \sim N(0, 1)$$

$$Q = X^T B X$$

$$B = B^2$$

Тогда $Q \sim \chi^2(r), r = \text{rank}(B = \text{tr}(B))$

$$Q = \sum_{k=1}^n (u_k^T X)^2 \sim \chi^2(r)$$

$$u_k^T \sim N(u_k^T E X, u_k^T I_n u_k) = N(0, 1)$$

$$\text{Cov}(u_k^T X, u_j^T X) = 0$$

$$B = U \Lambda U^T$$

$$\text{rank} B = \text{rank} \Lambda = \text{tr} \Lambda$$

$$\text{tr} B = \text{tr}(U \Lambda U^T)$$

$$\text{Заметим что } B_{j,j} = \lambda_j u_j u_j^T = \lambda_j$$

Теорема Фишера

$$X_1, \dots, X_n \sim N(\mu, \sigma^2) \Rightarrow$$

$$1. \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$2. \frac{nS_*^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$3. S^2, \bar{X} - \text{нез}$$

$$Y_j = \frac{X_j - M}{\sigma}$$

$$\bar{X} \frac{1}{\sigma} (\bar{X} - M)$$

$$S_*^2(Y) = \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2 = \frac{S_*^2(X)}{\sigma^2}$$

$$\bar{Y} = \frac{\sum Y_j}{n} = \frac{(\frac{1}{n}, \dots, \frac{1}{n})}{b} (Y_1, \dots, Y_n)^T = bY$$

$$nS_*^2(Y) = (Y - BY)^T (Y - BY) = Y^T (I - B)^T (I - B) Y \sum \Xi^2(\text{tr}(I - B))$$

Для того чтобы доказать третье утверждение

$$b(I - B) = b - b = 0$$

Тогда мы пользуемся первой леммой

Таким образом теорема Фишера доказана.

Мы доказали теорему Фишера давайте теперь с помощью теоремы мы рассмотрим задачу построения доверительных интервалов нормального закона

- σ^2 - известно, $\mu = ?$

$$\text{Рассмотрим два варианта: } \frac{X_1 - \mu}{\sigma} \sim N(0, 1)$$

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$$

$$\text{Доверительный интервал уровня } 1 - \alpha \left[\bar{X} - \frac{q_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}, \bar{X} + \frac{q_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \right]$$

- μ - известно, $\sigma^2 = ?$

$$-q \leq \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} \leq q$$

$$-q\sigma \leq \sqrt{n}(\bar{X} - \mu) \leq q\sigma$$

$$\frac{\sqrt{n}(\bar{X} - \mu)}{q} \leq \sigma$$

$$-\frac{\sqrt{n}(\bar{X} - \mu)}{q} \leq \sigma$$

Рассмотрим следующую статистику:

$$\sum \frac{(X_i - \mu)^2}{\sigma^2} \sim \Xi^2(n)$$

$$P(q_{\frac{\alpha}{2}} \leq \sum \frac{(X_i - \mu)^2}{\sigma^2} \leq q_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Доверительный интервал:

$$\frac{\sum (X_i - \mu)^2}{q_{1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{\sum (X_i - \mu)^2}{q_{\frac{\alpha}{2}}}$$

Давайте теперь рассмотрим задачу построения доверительного интервала

$$\mu = ?, \sigma^2 = ?$$

Воспользуемся теоремой Фишера:

$$\frac{nS_*^2}{\sigma^2} \sim \Xi^2(n-1)$$

$$\text{Доверительный интервал: } \frac{nS_*^2}{q_{1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{nS_*^2}{q_{\frac{\alpha}{2}}}$$

Определение. Распределение Стьюдента

X_0, X_1, \dots, X_n - нез, $N(0, 1)$

$$\frac{X_0}{\sqrt{\frac{1}{n} \sum_{k=1}^n X_k^2}} \sim T(n)$$

$$\sqrt{\frac{1}{n} \sum_{k=1}^n X_k^2}$$

n - степени свободы (deg of freedom)

Давайте выведем статистику:

$$\frac{\sqrt{n} \frac{\bar{X} - \mu}{\sigma}}{\sqrt{\frac{1}{n-1} \frac{ns_*^2}{\sigma^2}}} = \sqrt{n-1} \frac{\bar{X} - \mu}{S_*} \sim T(n-1)$$

$$\frac{\sqrt{n} \frac{\bar{X} - \mu}{\sigma}}{\sqrt{\frac{1}{n-1} \frac{(n-1)S^2}{\sigma^2}}} = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

Доверительный интервал:

$$\bar{X} - \frac{q_{1-\frac{\alpha}{2}} S}{\sqrt{n}}, \bar{X} + \frac{q_{1-\frac{\alpha}{2}} S}{\sqrt{n}}$$

Определение. Распределение Фишера

$$\Xi_n^2 \sim \Xi^2(n)$$

$$\Xi_m^2 \sim \Xi^2(m)$$

Они независимы

$$\frac{\Xi_n^2(n)}{\Xi_m^2(m)} \sim F(n, m)$$

$$\Xi_m^2(m)$$

15 Асимптотические доверительные интервалы

Раньше мы говорили $P(\theta \in (l_n, r_n)) \geq 1 - \alpha$

Теперь же мы будем говорить $\lim_{n \rightarrow \infty} P(\theta \in (l_n, r_n)) \geq 1 - \alpha$

$T(X, \theta) \xrightarrow{d} G$ не зависит от θ

1. ЦПТ и ее следствия

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \rightarrow N(0, 1)$$

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \rightarrow N(0, 1)$$

$$\text{Доверительный интервал: } \bar{X} \pm \frac{q_{1-\frac{\alpha}{2}} S}{\sqrt{n}}$$

$$\sqrt{n} \frac{S_*^2 - \sigma^2}{\sqrt{\widehat{\beta}_4 - S_*^4}} \rightarrow N(0, 1)$$

$$\text{Доверительный интервал: } S_*^2 \pm \frac{q_{1-\frac{\alpha}{2}} \sqrt{\widehat{\beta}_4 - S_*^4}}{\sqrt{n}}$$

2. Теорема об асимптотике среднего члена вариационного ряда

$$\sqrt{n} \frac{X_{(\lfloor np \rfloor)} - q_p}{\sqrt{p(1-p)}} f(q_p) \rightarrow N(0, 1)$$

$$\text{Доверительный интервал для медианы: } p = \frac{1}{2}$$

$$\sqrt{n} f(q_p) \frac{X_{(\lfloor \frac{n}{2} \rfloor)} - q_p}{\frac{1}{2}} \quad (\text{зачастую } f \text{ это константа})$$

$$\text{Доверительный интервал: } X_{(\lfloor \frac{n}{2} \rfloor)} \pm \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n} \cdot \text{const}}$$

3. Доверительные интервалы из асимптотической нормальности оценок максимального правдоподобия

4. Теорема об асимптотике крайнего члена вариационного ряда

$$n(1 - F(X_{m+1-s})) \rightarrow \Gamma(s, 1)$$

Пример:

$$U[0, \theta]$$

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \leq x \leq \theta \\ 1, & x > \theta \end{cases}$$

$$\frac{nX_{(r)}}{\theta} \rightarrow \Gamma(r, 1)$$

$$q_{\frac{\alpha}{2}} \leq \frac{nX_{(r)}}{\theta} \leq q_{1-\frac{\alpha}{2}} \Rightarrow \frac{nX_{(r)}}{q_{1-\frac{\alpha}{2}}} \leq \theta \leq \frac{nX_{(r)}}{q_{\frac{\alpha}{2}}}$$

16 Проверка статистических гипотез

Нам надо будет выделить основное предположение (по умолчанию) и альтернативное предположение (наше подозрение или то, что мы хотим доказать)

Давайте рассуждать:

рациональное, с точки зрения инопланетянина (не опираться на жизненный опыт)

В стране Н с континентальной системой права и есть уголовный суд. Какое будет основное предположение для судьи? Не виновен. А альтернативное? Виновен.

Давайте посмотрим на другой пример

В паспорте у некоторых написана буква М, а у других Ж, посмотрим связаны ли буквы и успеваемость? По умолчанию не связаны, а альтернативное: М и Ж учатся по разному. Пусть мы сидим на филфаке, тогда есть мнение, что девочки учатся лучше, или же мы сидим на программировании и тогда мальчики учатся лучше. Альтернатива не всегда является отрицанием первого.

Еще один пример:

Пусть робот кидает монетку. По умолчанию робот кидает честно, альтернативно робот жулик или он жулик в определенную сторону.

Замеры показателя (температуры человека). По умолчанию 36.6, альтернативно мы можем подозревать, что температура $\neq 36.6$ или же можем рассмотреть перепад в одну из сторон (в зависимости от болезни)

Пусть есть фактор цена на недвижимость и есть фактор расстояние до центра города. Основное предположение: они не зависят, Альтернатива: ближе к центру - больше цена.

Мы можем изготовить некое вещество и посмотрим как оно влияет на здоровье. Основное: не влияет, альтернатива: it depends.

X_1, \dots, X_n - выборка в широком смысле.

$(X_1, \dots, X_n) \sim F$

H_0 - нулевая гипотеза.

H_1 - альтернатива.

Так же пусть нам дали уровень значимости

$\alpha \in (0, 1)$ (по умолчанию 0.1, 0.05, 0.01, 0.001)

Статистический тест (критерий)

$$\delta(X, \alpha, H_0, H_1) = \begin{cases} \text{accept } H_0 \\ \text{reject } H_0 (w. \text{ respect to } H_1) \end{cases} \quad \text{То есть в первом случае}$$

данные противоречат H_0 , а втором противоречат.

Но это не значит, что мы доказали утверждение.

Пусть у нас есть функция $T(X)$ - статистика критерия

$T(X)$ либо в точности, либо в пределе стремится к G при условии $H_0(\sim$

or \rightarrow)
 $P(T(X) \in T_0(\alpha) | H_0) = 1 - \alpha$
 if $T(x) \in T_1(\alpha)$: reject H_0
 else: accept H_0
 $T_0(\alpha)$ - область принятия
 $T_1(\alpha)$ - область опровержения
 Далее на лекции идет пример с левосторонним, правосторонним и двойным тестом

1. left: $T_0(\alpha) = [q_\alpha, +\infty)$
2. right: $T_1(\alpha) = (-\infty, q_\alpha)$
3. two: $T_0(\alpha) = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}], T_1(\alpha) = \overline{T_0(\alpha)}$

$p_l = P(U \leq T(x) | H_0)$
 $p_r = P(U > T(x) | H_0)$
 $p = 2\min(p_l, p_r)$
 if $p < \alpha$: reject H_0
 else: accept H_1

17 Статистические критерии и доверительные интервалы

Когда мы строили доверительные интервалы то мы зажимали статистику между квантилями. Это похоже на двухсторонний тест.

$X_1, \dots, X_n \sim F_\theta$
 $T(X, \theta) \rightarrow U \sim G$
 $P(q_{\frac{\alpha}{2}} \leq T(X, \theta) \leq q_{1-\frac{\alpha}{2}}) = 1 - \alpha$
 $H_0 : \theta = \theta_0$
 $P(T(X, \theta_0) \in T_0(\alpha) | \theta = \theta_0) = 1 - \alpha$
 $H_1 = \theta \neq \theta_0, \theta > \theta_0, \theta < \theta_0$

Пример:

1) $X_1, \dots, X_n \sim F, \mu = EX_1, \exists \text{Var } X$
 $H_0 : \mu = \mu_0$

$T(X) = \sqrt{n} \frac{\bar{X} - \mu_0}{S} \rightarrow N(0, 1)$ если $\mu = \mu_0$

При $H_1 : \mu \neq \mu_0$ у нас двухсторонняя критическая область

При $H_1 : \mu > \mu_0$ у нас правосторонняя критическая область

При $H_1 : \mu < \mu_0$ у нас левосторонняя критическая область

Далее для этого приводится пример с больницей (нам либо надо просто проверить, что температура не стандартная, либо нам важно знать, что она больше нормы, либо нам важно знать, что она ниже нормы)

$$P(\sqrt{n} \frac{\bar{X} - \mu_0}{S} \in T_0(\alpha) | \mu \neq \mu_0) = P(\sqrt{n} \frac{\bar{X} - \mu}{S} + \frac{\mu - \mu_0}{S} \sqrt{n} | \mu \neq \mu_0)$$

Это будет стремиться либо к $\Phi(-\infty)$ либо $1 - \Phi(+\infty)$

18 Критерий Колмогорова

$$X_1, \dots, X_n \sim F$$

$$H_0 : F = F_0 \text{ (} F_0 \text{ - непр)}$$

$$H_1 : F \neq F_0$$

Идея основана на теореме Колмогорова (было в начале семестра)

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|, F_n - \text{эмпирическая функция распределения}$$

if $D_n > q_{1-\alpha}$ then reject H_0 else accept H_0

1) $n \geq 20$ работает хорошо, при маленьких n есть спец таблицы

2) Так же есть приближенные формулы для D_n

$$3) H_0 : F = F(\theta), H_1 : \neg H_0 \Rightarrow D_n = \sqrt{n} \sup_x |F_n(x) - F_0(x, \theta)|$$

$$\theta \rightarrow \hat{\theta}$$

В пределе будет более сложная формула

19 Критерий Смирнова

$$X_1, \dots, X_n$$

$$Y_1, \dots, Y_m$$

Они независимы

$$H_0 : F_X = F_Y (= F_0)$$

$$H_1 : \neq H_0$$

Тут идея основана на формуле Смирнова

$$D_{n,m} = \sqrt{n} \sup_x |F_n(x) - F_m(x)|$$

$$T_1(\alpha) = (q_{1-\alpha}, +\infty)$$

20 Критерий типа хи-квадрат

20.1 Критерий согласия Пирсона

$$X_1, \dots, X_n \sim F(x) - \text{непрерывная}$$

Давайте дискретизируем данные

$$\Delta_1 : \nu_1 - \text{количество элементов выборки попадающих } \Delta_1$$

...

$$\Delta_N$$

$$p_{\Delta_k} = \int_{\delta_k} p(x) dx, p(x) = F'(x)$$

Рассмотрим $\{1, 2, \dots, N\}$, $p = (p_1, \dots, p_N)$ - настоящий вектор вероятностей

$p_0 = (p_{01}, \dots, p_{0N})$ - ожидаемый фиксированный вектор вероятностей

ν_k - количество элементов в выборке типа k

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

$$n = \sum_{k=1}^N \nu_k$$

$$\Xi_N^2 = \sum_{k=1}^N \frac{(\nu_k - np_{0k})^2}{np_{0k}}$$

Теорема

$$\Xi_N^2 \xrightarrow{n \rightarrow \infty} \Xi^2(N-1) \text{ при условии } H_0$$

Доказательство

$N = 2 :$

$$\frac{(\nu_1 - np_{01})^2}{np_{01}} + \frac{(\nu_2 - np_{02})^2}{np_{02}} = \frac{(\nu_1 - np_{01})^2}{np_{01}} + \frac{(n - \nu_1 - n(1 - p_{01}))^2}{n(1 - p_{01})} = \frac{(\nu_1 - np_{01})^2}{n} \left(\frac{1}{p_{01}} + \frac{1}{1 - p_{01}} \right) = \frac{(\nu_1 - np_{01})^2}{np_{01}(1 - p_{01})}$$

Без квадрата по ЦПТ это стремится к $N(0, 1)$, но мы можем навесить непрерывную функцию возведения в квадрат и получим то, что нам нужно

Для классического критерия хи-квадрат у нас правосторонняя критическая область (потому что в сумме при H_0 будет маленькая разность в квадрате и тд и тп очев крч) **Замечание**

Критерий состоятельный

Пусть у нас есть некоторые типы семян (круглые и желтые, морщинистые и желтые, круглые и зеленые, морщинистые и зеленые)

$$\nu = (315, 101, 108, 32)$$

$$p_0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right)$$

$$\Xi_4^2 = 0.47$$

$$df = 3$$

$$p_value = 1 - (df(0.47)) = 0.97$$

Замечание

$$n \geq 50, \nu_j \geq 5 \text{ **Замечание**}$$

Оптимальные способы дискретизации существуют

Замечание

Для проверки согласованности с нормальным законом следует использовать спец тесты (Шапиро, коэффициент эксцесса)

$$H_0 : p = p_0(\theta), \theta \in \Theta \subset \mathbb{R}^d, d < N - 1$$

$$\Xi_N^2 = \sum_{k=1}^N \frac{(\nu_k - np_{0k}(\theta))^2}{np_{0k}(\theta)} \Big|_{\theta=\hat{\theta}}$$

Теорема

$$p_0(\theta) > 0 \forall \theta$$

$$\frac{\partial p_0}{\partial \theta}, \frac{\partial^2 p_0}{(\partial \theta)^2} - \text{непрерывная}$$

$$\text{Тогда } \Xi_N^2 \rightarrow X^2(N - 1 - d)$$

$$rk\left(\frac{\partial p_{0k}}{\partial \theta_j}\right)_{1 \leq k \leq N, 1 \leq j \leq d} = d$$

20.2 Критерий однородности

Пусть у нас K независимых выборок все они из $\{1, 2, \dots, N\}$

Пусть $p^{(j)}$ - истинный вектор вероятностей для соответствующей выборки

$$H_0 : p^{(1)} = \dots = p^{(k)}$$

$$H_1 : \neq H_0$$

ν_{ij} - количество элементов типа j в i -ой выборке

$$n_i = \sum_j \nu_{ij} = \nu_{i*} - \text{объем } i\text{-ой выборки}$$

$$n = n_1 + \dots + n_k$$

Пусть $p^{(1)}, \dots, p^{(k)}$ - известны

$$\Xi_{n_1}^2 = \sum_{j=1}^N \frac{(\nu_{ij} - n_i p_j^{(i)})^2}{n_i p_j^{(i)}}, df = N - 1$$

$$\Xi_{n_1, n_2, \dots, n_k}^2 = \sum_{i=1}^K \Xi_{n_i}^2, df = k(N - 1)$$

Рассмотрим $p^{(1)}, \dots, p^{(k)}$ - не известны

$$df = k(N - 1) - (N - 1) = (N - 1)(k - 1)$$

$$\hat{p}_j = \frac{\nu_{1j} + \dots + \nu_{kj}}{n} = \frac{\nu_{*j}}{n}$$

$$L(\dots) = p_1^{(\nu_{*1})} \dots p_k^{(\nu_{*k})}, n = \sum_j \nu_{*j}$$

TODO пример 2x2 10 лекция 55:40

20.3 Критерий независимости

$$X_1, \dots, X_n : \{1, 2, \dots, N\}$$

$$Y_1, \dots, Y_n : \{1, 2, \dots, M\}$$

ν_{ij} - количество пар, в которых первая компонента равна i , а вторая j

Это можно представить в виде таблицы сопряженности

Просуммируем по каждому столбцу и по каждой строчке

$$\text{Пусть } p_{ij} = P(X = i, Y = j)$$

$$p_{xi} = P(X = i)$$

$$p_{yj} = P(Y = j)$$

$$H_0 : p_{ij} = p_{xi} p_{yj} \forall i, j$$

$$H_1 : \neg H_0$$

$$\Xi^2 = \sum_{1 \leq i \leq N, 1 \leq j \leq M} \frac{(\nu_{ij} - p_{ij} n)^2}{n p_{ij}}, df = MN - 1$$

$$df = MN - 1 - (N - 1) - (M - 1) = MN - N - M + 1 = N(M - 1) - (M - 1) = (M - 1)(N - 1)$$

$$\widehat{p_{Xi}} = \frac{\nu_{i*}}{n}$$

$$\widehat{p_{Yj}} = \frac{\nu_{j*}}{n}$$