Математическая статистика

Матвеев Сергей M3338

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.d.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 фукнция распределения:
$\mu_n(t) = \sum\limits_{i=1}^{n} \mathbb{1}(X_i \le t)$
$F_n(t) = \dfrac{\mu_n(t)}{n}$ - эмпирическая функция распределения.

Замечение: Описывает эмпирическое распределениею.

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

$\mathbb{1}(\mathbb{X}_{\mathbb{i}} \leq \mathbb{t}) \sim Bern(F(t))$

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

$\operatorname{Var}(F_n(t)) = \dfrac{F(t)(1 - F(t))}{n}$

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

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

$= \sqrt{n}\dfrac{F_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}}$ (ассимптотическая нормальность)

Теорема Гливенко-Кантелли

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

Теорема Колмогорова

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

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

Такая функция называется функцией Колмогорова

Теорема Смирнова

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

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

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

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

3 Выборочные моменты

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

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

$\overline{g(X)} = \dfrac{1}{n}\sum\limits_{k=1}^{n} g(X_k)$, $g : \mathbb{R} \to \mathbb{R}$

$\widehat{\alpha_k} = \overline{X^k} = \dfrac{1}{n}\sum\limits_{j=1}^{n} X_j^k$ - к-ый выборочный момент.

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

$\operatorname{Var}\widehat{\alpha_k} = \dfrac{1}{n}\operatorname{Var}(X_1^k) = \dfrac{1}{n}(EX_1^{2k} - (EX_1^k)^2)$

По ЦПТ получаем:

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

$\sqrt{n}\dfrac{\widehat{\alpha_k}-\alpha_k}{\widehat{\alpha_{2k}}-\widehat{\alpha_k}^2} = \sqrt{n}\dfrac{\widehat{\alpha_k}-\alpha_k}{\alpha_{2k}-\alpha_k^2}\cdot\dfrac{\alpha_{2k}-\alpha_k^2}{\widehat{\alpha_{2k}}-\widehat{\alpha_k}}$ - первый множитель по ЦПТ сходится к $N(0,1)$ по распределнию
Давайте посмотрим что будет со вторым множителем. Он будет сходиться к 1 по вероятности.
Таким обрбазом:
$\sqrt{n}\dfrac{\widehat{\alpha_k}-\alpha_k}{\alpha_{2k}-\alpha_k^2}\cdot\dfrac{\alpha_{2k}-\alpha_k^2}{\widehat{\alpha_{2k}}-\widehat{\alpha_k}} \xrightarrow{d} N(0,1)$
А почему вторая дробь сходится к единице?
$\widehat{\alpha_k} \xrightarrow{P} \alpha_k$ (по ЗБЦ, это называется состоятельность)
$\widehat{\alpha_{2k}}-\widehat{\alpha_k}^2 \xrightarrow{P} \alpha_{2k}-\alpha_k^2$
$\overline{X}$ - выборочное среднее.
$\widehat{\beta_k} = \overline{(X-\overline{X})^k} = \dfrac{1}{n}\sum\limits_{j=1}^{n}(X_j-\overline{X})^k$ - к-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.
$S_* = \overline{X^2} - (\overline{X}^2)$
$\widehat{\beta_k} = Poly(\widehat{\alpha_k},\ldots,\widehat{\alpha_1})$
$\widehat{\alpha_1},\ldots,\widehat{\alpha_k}$ - 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
$(\xi_n-\mu)\sqrt{n}\cdot\dfrac{1}{\sqrt{n}} \xrightarrow{d} 0$ (Первое сходится к нормальному многомерному закону, а второе к нулю)
Мы знаем, что для вырожденных величин сходимость по распределению и вероятности равносильны.
Тогда мы можем написать $(\xi_n-\mu)\sqrt{n}\cdot\dfrac{1}{\sqrt{n}} \xrightarrow{P} 0$
$\sqrt{n}(\varphi(\xi_n)-\varphi(\mu)) \to$??
$\varphi \in C^1(\mathbb{R}^m)$, $\varphi : \mathbb{R}^m \to \mathbb{R}$
По Тейлору:
$\varphi(\xi_n) \approx \varphi(\mu) + \nabla\varphi(\widetilde{\mu})(\xi_n-\mu)$ (Остаток в форме Лагранжа)
$\nabla\varphi(\widetilde{\mu}) \xrightarrow[n\to\infty]{} \nabla\varphi(\mu)$
$\varphi(\xi_n)-\varphi(\mu) \approx \nabla\varphi(\mu)(\xi_n-\mu)$

$\mathrm{Var}(\dots) \approx \mathrm{Var}(\nabla\varphi(\mu)(\xi_n - \mu)) = \mathrm{Var}(\nabla\varphi(\mu)\xi_n) = \nabla\varphi \,\mathrm{Var}\, \xi_n (\nabla\xi_n)^T$
Вернемся к следующему равентву:
$\varphi(\xi_n) - \varphi(\mu) \approx \nabla\varphi(\mu)(\xi_n - \mu)$
$\sqrt{n}(\varphi(\xi_n) - \varphi(\mu)) \approx \sqrt{n}\nabla\varphi(\mu)(\xi_n - \mu)$
$\sqrt{n}(\xi_n - \mu) \to N(0, \Sigma)$ (Как мы и говорили выше)
А мы домножаем вектор на градиент, поэтому:
$\sqrt{n}\nabla\varphi(\mu)(\xi_n - \mu) \to N(0, \nabla\varphi(\mu)\Sigma(\nabla\varphi(\mu))^T)$
Это называется дельта метод.
Теорема
Пусть $\xi_n = (\overline{X}, \dots, \overline{X^k})$
Многомерная версия ЦПТ
$\sqrt{n}(\xi_n - \alpha) \xrightarrow{d} N(0, \Sigma)$
$\alpha = (\alpha_1, \dots, \alpha_k)$
$\Sigma = \mathrm{Var}(X_1, \dots, X_1^k)$
$\varphi : \mathbb{R}^k \to \mathbb{R}$
$\varphi \in C^1(\mathbb{R}^k)$
$\sigma^2 = \nabla\varphi(\alpha)(\nabla\varphi(\alpha))^T > 0$
Тогда (продолжение теоремы):
$\sqrt{n}\dfrac{\varphi(\xi_n) - \varphi(\alpha)}{\sigma} \xrightarrow{d} N(0,1)$
Кроме того:
$\sigma = \sigma(\alpha) \in C^1(\mathbb{R}^k) \Rightarrow \sqrt{n}\dfrac{\varphi(\xi_n) - \varphi(\alpha)}{\sigma(\xi_n)} \xrightarrow{d} N(0,1)$
Коэффициент аассиметрии: $\dfrac{E(X - EX)^3}{\sigma^3} = \gamma$
$\dfrac{\widehat{\beta_3}}{S_*^3} = \widehat{\gamma}$
Коэффициент эксцесса: $\dfrac{E(X - EX)^4}{\sigma^4} - 3$
$\dfrac{\widehat{\beta_4}}{S_*^4} - 3$
Пусть у нас есть две выборки:
$X_1, \dots, X_n$
$Y_1, \dots, Y_n$
$\mathrm{Cov}(X, Y) = EXY - EX \cdot EY = E(X - EX)(Y - EY)$
$S_{*XY} = \dfrac{1}{n}\sum_j (X_j - \overline{X})(Y_j - \overline{Y}) = \dfrac{1}{n}\sum_j X_j Y_j - \overline{XY}$
$\rho = \dfrac{\mathrm{Cov}(X, Y)}{\sqrt{VarX \cdot VarY}}$
$\rho = \dfrac{S_{*XY}}{S_{*X} \cdot S_{*Y}}$ - Выборочный коэффициент корреляции (коэффициент корреляции Пирсона)

5 Порядковые статистики

Определение. Вариационный ряд
$X_{(1)} \leq X_{(3)} \leq \dots \leq X_{(n)}$ - вариационный ряд
Определение. Порядковая статистика
$X_{(k)}$ - к-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
$\alpha \in (0, 1)$
$\exists 0 \leq k \leq n - 1 : \dfrac{k}{n} \leq \alpha < \dfrac{k+1}{n}$
$X_{(k+1)}$ - выборочный квантиль порядка α
$\alpha = 0$ $\min(X)$ - нулевой квартиль
$F^{-1} = \sum\{x \in \mathbb{R} F_n(x) \leq \alpha\}$
$\alpha = \dfrac{1}{4}$ - первый вартиль (нижний квартиль)
$\alpha = \dfrac{1}{2}$ - второй квартиль (выборочная медиана)
$\alpha = \dfrac{3}{4}$ - третий квартиль (верхний квартиль)
$\alpha = 1$ - $\max(X)$ (четвертый квартиль)
$n = 2m \Rightarrow = med(X) = \dfrac{X_(m) + X_(m+1)}{2}$
$n = 2m + 1 \Rightarrow med(X) = X_(m+1)$
$IQR = \Delta$ между верхним и нижним квартилем.
$P(X_{(k)} \leq t) = P(\mu_n(t) \geq k) = \sum_{j=k}^{n} C_n^j F^j(t)(1 - F(t))^{n-j}$
$B(z, a, b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z t^{a-1}(1-t)^{b-1} dt = B(F(t), k, n-k+1)$
$0 \leq z \leq 1$
Пусть $p()$ - теоретическая плотность, то есть $p = F'$
$(P(X_(k) \leq t))'_t = \dfrac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \cdot F^{k-1}(t)(1 - F(t))^{n-k} \cdot p(t)$ - плотность к-й порядковой статистики
Рассуждая аналогично можно получить плотность для двухмерного случая:
$g(x_1, x_2) = \dfrac{n!}{(k-1)!(r-k-1)!(n-k)!} \cdot F^{k-1}(x_1)(F(x_2) - F(x_1))^{r-k-1}(1 - F(x_2))^{n=r} p(x_1) p(x_2)$ - 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: $\dfrac{K(n)}{n} \to const \in (0,1)$
Крайний член вариационного ряда:
$X_{(r)}$, r - огр.
$X_{(n+1-s)}$, s - огр.
Теорема об ассимптотике среднего члена вариационного ряда
$0 < \alpha < 1$ - теоретическая плотность.
q_α - теоретический квантиль порядка α
$p \in C^1($ окр-сть $q_\alpha)$
$p(q_\alpha) > 0$
Тогда:
$\sqrt{n} \cdot f(q_\alpha) \dfrac{X_{(\lfloor n\alpha \rfloor)} - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} N(0,1)$
Идея доказательства
Пусть $\lfloor n\alpha \rfloor = k$
Мы умеем писать плотность для $X_{(k)}$
Затем у нас идет преобразование:
$g(x) = \sqrt{n} p(q_\alpha) \dfrac{x - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \rightsquigarrow p_{g(X_{(k)})}(t) = p_{X_{(k)}}(g^{-1}(t)) |g^{-1}(t)'_t|$ (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 $q_\alpha)$
$p(q_\alpha) > 0$
Тогда в пределе наша новая плотность будет стремиться к плотности нормального стандартного закона.
Пример. Распределение Коши.
$Couchy(\mu, 1)$ $\alpha = \dfrac{1}{2}$
$2f(\mu)\sqrt{n}(X_{(\lfloor \frac{n}{2} \rfloor)}) = \dfrac{1}{\pi} \cdot \dfrac{1}{1 + (\mu - \mu)^2} \cdot 2\sqrt{n}(X_{(\lfloor \frac{n}{2} \rfloor)} - \mu)$
Теорема об ассимптотике 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 оценит θ в виде $\widehat{\theta} = \widehat{\theta}(X_1, \dots, X_n)$
Замечание:

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

Определение. Состоятельность
$\widehat{\theta}$ - состоятельная оценка $\theta \Leftrightarrow \widehat{\theta} \xrightarrow{P} \theta$
Определение. Несмещенность
$b.as(\widehat{\theta}) \stackrel{def}{=} E\widehat{\theta} - \theta$ - смещение
$b.as(\widehat{\theta}) = 0 \Leftrightarrow$ несмещенная
Определение. Ассимптотическая нормальность
$\sqrt{n}(\widehat{\theta} - \theta) \to N(0, \Sigma_\theta)$
Пример

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

Определение. Эффективность (оптимальность)
$\widehat{\theta}_1$ эффективнее $\widehat{\theta}_2 \Leftrightarrow MSE\widehat{\theta}_1 < MSE\widehat{\theta}_2$
$MSE\widehat{\theta} = E\left\|\widehat{\theta} - \theta\right\|^2 = E(\widehat{\theta} - \theta)^T(\widehat{\theta} - \theta)$
Утверждение
$MSE\widehat{\theta} = tr(\operatorname{Var}\widehat{\theta}) + \left\|b.as\widehat{\theta}\right\|^2$
Доказательство
$MSE = E(\widehat{\theta} - \theta)^T(\widehat{\theta} - \theta) = E(\widehat{\theta} - E\widehat{\theta} + E\widehat{\theta} - \theta)^T(\widehat{\theta} - E\widehat{\theta} + E\widehat{\theta} - \theta) = E(\widehat{\theta} - E\widehat{\theta})^T(\widehat{\theta} - E\widehat{\theta}) = \sum \operatorname{Var}\widehat{\theta}_i + \left\|b.as\widehat{\theta}\right\|^2$

1. Ассимптотическая нормальность $\Rightarrow$ состоятельность
   $\widehat{\theta} - \theta = \dfrac{1}{\sqrt{n}}(\widehat{\theta} - \theta) \xrightarrow{P} 0$

2. Ассимптотическая нормальность $\Rightarrow b.as\widehat{\theta} \to 0$
   Пусть $d = 1$
   $P(|\theta - E\widehat{\theta}| > \varepsilon) = P(\dfrac{\sqrt{n}|\theta - E\widehat{\theta}|}{\sigma} > \dfrac{\varepsilon\sqrt{n}}{\sigma}) = 1 - P(\dots < \dfrac{\varepsilon\sqrt{n}}{\sigma}) \approx 1 - (2\Phi(\dfrac{\varepsilon\sqrt{n}}{\sigma}) - 1) = 2(1 - \Phi(\dfrac{\varepsilon\sqrt{n}}{\sigma})) \to 0$

3. Состоятельность $\Rightarrow b.as\widehat{\theta} \to 0$
   Следует из усиленного закона больших чисел
   $\overline{X} \xrightarrow{a.s} \mu \Rightarrow E\overline{X} \to \mu$ (По теореме Лебега о мажорируемой сходимости)

4. Пусть $d = 1$, $b.as\widehat{\theta} \to 0, \operatorname{Var}\widehat{\theta} \to 0 \Rightarrow \widehat{\theta}$ - сост.

7 Метод моментов

Рассмотрим $g_1, \dots, g_d$
$\exists E g_1(X_1) = m_1(\theta_1, \dots, \theta_d)$
$\exists E g_2(X_2) = m_2(\theta_1, \dots, \theta_d)$
$\dots$
$\exists E g_d(X_d) = m_d(\theta_1, \dots, \theta_d)$
$$\begin{cases} \overline{g_1(X)} = m_1(\widehat{\theta_1}, \dots, \widehat{\theta_d}) \\ \dots \\ \overline{g_d(X)} = m_d(\widehat{\theta_1}, \dots, \widehat{\theta_d}) \end{cases}$$
Пусть $\exists!$ решение:
$$\begin{cases} \widehat{\theta_1} = \alpha_1(\overline{g_1(X)}, \dots, \overline{g_d(X)}) \\ \dots \\ \widehat{\theta_d} = \alpha_d(\overline{g_1(X)}, \dots, \overline{g_d(X)}) \end{cases}$$
Тогда это будет оценка методов моментов.
По умолчанию $g_j(x) = x^j$
Пример
$U[\theta_1, \theta_2]$, $\theta_1 < \theta_2$
$g_1(x) = x, g_2(x) = x^2$
$E g_1(X_1) = E X_1 = \dfrac{\theta_1 + \theta_2}{2}$
$E g_2(X_1) = E X_1 = \dfrac{(\theta_2 - \theta_1)^2}{12} + \dfrac{(\theta_1 + \theta_2)^2}{4} = \dfrac{\theta_2^2 - 2\theta_1\theta_2 + \theta_1^2}{12} + \dfrac{\theta_1^2 + 2\theta_1\theta_2 + \theta_2^2}{4} = \dfrac{\theta_1^2 + \theta_1\theta_2 + \theta_2^2}{3}$
Далее составим уравнения:

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

$3\overline{X^2} = \widehat{\theta}_1^2 + \widehat{\theta_1}(2\overline{X} - \widehat{\theta_1}) + (2\overline{X} - \widehat{\theta_1})^2$
$\widehat{\theta}_1^2 - 2\overline{X}\widehat{\theta_1} + 4(\overline{X}^2 - 3\overline{X^2}) = 0$
Считаем дискриминант деленный на четыре:
$\dfrac{D}{4} = (\overline{X})^2 - 4(\overline{X})^2 + 3\overline{X^2} = 3(\overline{X^2 - (\overline{X})^2}) = 3S_*^2$
У нас будет два случая:

1. $\begin{cases} \widehat{\theta_1} = \overline{X} + \sqrt{3}S_* \\ \widehat{\theta_2} = \overline{X} - \sqrt{3}S_* \end{cases}$

2. $\begin{cases} \widehat{\theta_1} = \overline{X} - \sqrt{3}S_* \\ \widehat{\theta_2} = \overline{X} + \sqrt{3}S_* \end{cases}$

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

1. если $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$ - состоятельаня оценка
2. если $(\overline{g_1(X)}, \dots, \overline{g_d(X)})$ - ассимптотически нормальные и $g_1, \dots, g_d$ - гладкие, то каждая из оценок ассимтотичесик 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
$\theta^* = \underset{\widehat{\theta}}{argmax}(L(x, \theta))$ - 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 $\ln L(x|\theta)$; $\dfrac{\partial \ln L(x, \theta)}{\partial \theta}$
3. $\dfrac{\partial \ln L(x, \theta)}{\partial \theta} = 0$
4. Проверить достаточные условия максимума

Пример

1. $N(\theta_1, \theta_2)$
   $L(x, \theta) = \prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp{(-\frac{(x_i - \theta_1)^2}{2\theta_2})}$
   $\ln L(x, \theta) = \sum\limits_{i=1}^{n} [-\frac{1}{2} \ln 2\pi - \frac{1}{2}\theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2}]$
   $\frac{\partial \ln L(x, \theta_1)}{\theta_1} = \sum\limits_{i=1}^{n} \frac{2(x_i - \theta_1)}{2\theta_2} = \sum\limits_{i=1}^{n} \frac{x_i - \theta_1}{\theta_2}$
   $\frac{\partial \ln L(x, \theta)}{\partial\theta_2} = \sum\limits_{i=1}^{n} [-\frac{1}{2\theta_2} + \frac{(x - i - \theta_1)^2}{2\theta_2^2}]$
   $\sum\limits_{i=1}^{n} \frac{(x_i - \widehat{\theta_1})}{\widehat{\theta_2}} = 0 \Rightarrow \widehat{\theta_1} = \overline{X}$
   $\sum\limits_{i=1}^{n} [-\frac{1}{2\widehat{\theta_2}} + \frac{(x_i - \widehat{\theta_1})^2}{2\widehat{\theta_2}^2}] = 0 \Rightarrow \widehat{\theta_2} = S_*^2$

2. $U[0, \theta] : L(X, \theta) = \frac{1}{\theta^n} \mathbb{1}(X_i \in [0, \theta], \forall 1 \leq i \leq n)$
   $X_1, \ldots, X_n : p(X_i, \theta)$
   $L(X_1, \ldots, X_n, \theta) = \prod\limits_{i=1}^{n} p(x_i, \theta)$
   $L(X, \theta) = \theta^{\Sigma X_i} (1 - \theta)^{n - \Sigma X_i}$
   Вернемся к нашему примеру:
   Если $\widehat{\theta} < \max(X)$, то $L(x, \widehat{\theta}) = 0$
   Чем меньше θ, тем больше $\frac{1}{\theta}$
   $\Rightarrow \widehat{\theta} = \max(X)$

$Poly(1, p) : p = (p_1, \ldots, p_m)$
Рассмотрим частоты:
ν_1 - кол-во наблюдений типа 1
$\ldots$
ν_m - кол-во наблюдений типа m
Суммируем и смотрим на функцию правдоподобия
$L(X, p) = p_1 \ldots p_m$
$\ln L(X, p) = \sum\limits_{j=1}^{m-1} \nu_j \ln p_j + \nu_m \ln{(1 - P_1 - \cdots - p_{m-1})}$
$\frac{\partial \ln L \ldots}{\partial p_j} = \frac{\nu_j}{p_j} - \frac{\nu_m}{1 - p_1 - \cdots - p_{m-1}} = 0$
$\sum$ уравнения: $\nu_j (1 - \widehat{p_1} - \cdots - \widehat{p_{m-1}}) = \widehat{p_j} \cdot \nu_m$
$\widehat{p_m}(n - \nu_m) = \nu_m (1 - \widehat{p_m})$
$\widehat{p_m} n - \widehat{p_m} \nu_m = \nu_m$

$\widehat{p_m} = \frac{\nu_m}{n}$
$\widehat{p_j} = \frac{\nu_j \widehat{p_m}}{\nu_m} = \frac{\nu_j}{n}$

9 Информация Фишера

$d = 1 : L(X, \theta) = \prod p(X_j, \theta)$
$\ln L(X, \theta) = \sum \ln p(x_j, \theta)$
$V(X, \theta) = \frac{\partial \ln L \ldots}{\partial \theta} = \sum \frac{\partial \ln p \ldots}{\partial \theta}$ - 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. $EV^2(X, \theta) < \infty$

$\int_X L(X, \theta) dX = 1 \xrightarrow{\frac{\partial}{\partial \theta}} \int_X \frac{\partial L(.)}{\partial \theta} dX = \int_X \frac{\frac{\partial L(\ldots)}{\partial \theta}}{L(\ldots)} \cdot L(\ldots) dX = \int_X V(X, \theta) L(X, \theta) dX = EV(X, \theta) = 0$
$I(\theta) = \mathrm{Var}(V(X_i, \theta)) = E(V^2(X_i, \theta))$ - информация Фишера для всей выборки
$V(X, \theta) = \sum_j \frac{\partial \ln p(x, \theta)}{\partial \theta} \Rightarrow \mathrm{Var}(V(X, \theta)) = n \cdot \mathrm{Var} \frac{\partial \ln p(x, \theta)}{\partial \theta}$
$i(\theta)$ - информация Фишера для 1 наблюдения
$i(\theta) = E(\frac{\partial \ln p(x_j, \theta)}{\partial \theta})^2$
$\frac{\partial}{\partial \theta} \int_{\mathbb{R}} \frac{\partial \ln P(x, \theta)}{\partial \theta} \cdot p(x, \theta) dx = \int_{\mathbb{R}} \frac{\partial^2 \ln p(x, \theta)}{\theta} dx + \int_{\mathbb{R}} \frac{\partial \ln \ldots}{\partial \theta} \frac{\partial p \ldots}{\partial \theta} dx = E \frac{\partial^2 \ln p(x, \theta)}{\partial \theta^2} + E(\frac{\partial \ln o(x, \theta)}{\partial \theta})^2 = 0$
$i(\theta) = E(\frac{\partial \ln p(x_j, \theta)}{\partial \theta})^2 = -E \frac{\partial^2 \ln p(x, \theta)}{\partial \theta^2}$
Произвольное d:
$i(\theta) = -(E \frac{\partial^2 \ln p(X, \theta)}{\partial \theta_i \partial \theta_j})_{1 \leq i,j \leq d}$
$I(\theta) = n i(\theta)$
$N(\theta_1, \theta_2)$

$p(x, \theta_1, \theta_2) = \dfrac{1}{\sqrt{2\pi\theta_2}} \cdot \exp(-\dfrac{(x-\theta_1)^2}{2\theta_2})$
$\ln p(x, \theta_1, \theta_2) = -\dfrac{1}{2}\ln 2\pi - \dfrac{1}{2}\ln\theta_2 - \dfrac{(x-\theta_1)^2}{2\theta_2}$
$\dfrac{\partial \ln p(x, \theta_1, \theta_2)}{\partial\theta_1} = \dfrac{x-\theta_1}{\theta_2}$
$\dfrac{\partial \dots}{\partial\theta_2} = -\dfrac{1}{2\theta_2} + \dfrac{(x-\theta)}{}$ TODO:
Теорема. Неравенство Рао-Крамера
Модель регулярная, d = 1
$\tau(\theta)$ - оцениваемая функция
$\tau \in C^1$ (как правило $\tau(\theta) = \theta$)
$\widehat{\tau(\theta)} = \theta \Rightarrow \operatorname{Var}\widehat{\tau(\theta)} \geq \dfrac{[\tau'(\theta)]^2}{ni(\theta)}$
$\tau'(\theta) = \displaystyle\int \widehat{\tau(\theta)} \frac{\partial L(X,\theta)}{\partial\theta} dX = \int \widehat{\tau(\theta)} V(X,\theta) dX - EV(X,\theta) \cdot E\widehat{\tau(\theta)} = \operatorname{Cov}(V(X,\theta), \widehat{\tau(\theta)})$
$\operatorname{Cov}^2(V(X,\theta), \widehat{\tau(\theta)}) \leq \operatorname{Var}(V(X,\theta)) \cdot \operatorname{Var}(\widehat{\tau(\theta)})$
Замечания

1. $E\widehat{\tau(\tau)} - \tau(\theta) = bias(\theta) \neq 0$
   $E\widehat{\tau(\theta)} = \tau(\theta) + bias(\theta)$
   $\operatorname{Var}\widehat{\tau(\theta)} \geq \dfrac{[\tau'(\theta) + bias'(\theta)]^2}{ni(\theta)}$

2.

Многомерный случай
$\tau(\theta) : \mathbb{R}^d \to \mathbb{R}$
$\tau \in C^1$
$E\widehat{\tau\theta} = \tau\theta \Rightarrow \operatorname{Var}\widehat{\tau(\theta)} \geq \dfrac{\nabla\tau(\theta) i^{-1}(\theta) \nabla^T \tau(\theta)}{n}$
Пример
$N(\theta_1, \theta_2)$
$\tau(\theta_1, \theta_2) = \theta_1$
$E\widehat{\theta_1} = \theta_1 \Rightarrow \operatorname{Var}\widehat{\theta_1} \geq \dfrac{(1,0)\begin{pmatrix}\theta_2 & 0 \\ 0 & 2\theta_2^2\end{pmatrix} \cdot (1,0)^T}{n} = \dfrac{\theta_2}{n}$
Свойства оценки

1. Если $\exists$ несмещенная оптимальная оценка в регулярном случае то она совпадает с оценкой максимальног правдоподобия (ОМП)
   $\tau(\theta) = \theta$
   $V(X,\theta) = \dfrac{1}{a(\theta)}(\widehat{\theta} - \theta)$

10 Состоятельность ОМП

Пусть θ_0 - реальный параметр $\Rightarrow p_{\theta_0}(L(X,\theta_0) > L(X,\theta)) \to 1$
$\frac{L(X,\theta)}{L(X,\theta_0)} < 1$
$\frac{1}{n}\sum \ln\frac{p(X_j,\theta)}{p(X_j,\theta)} < 0$
По ЗБЧ $\Rightarrow E_{\theta_0}\ln\frac{p(x_j,\theta)}{p(x_j,\theta_0)} \leq E_{\theta_0}[\frac{p(X_j,\theta)}{p(X_j,\theta_0)} - 1] = \int_X p(X,\theta)dX - \int p(X,\theta_0)dX = 0$
Давайте введем события:
$S_n = \{X : \ln L(X,\theta_0) > \ln L(X,\theta_0 - a)\} \cap \{X : \ln L(X,\theta_0) > \ln L(X_i\theta_0 + a)\}$
$P_{\theta_0}(S_n) \to 1$
$A_n = \{X : |\widehat{\theta} - \theta_0| < a\}$
$B_n = \{X : \frac{\partial \ln L(X,\theta)}{\partial\theta}|_{\theta=\widehat{\theta}} = 0\}$
$S_n \subset A_nB_n \subset A_n \Rightarrow P(A_n) \to 1$
Давайте поговорим про свойства метода максимального правдоподобия

1. Принцип инвариантности
   $\theta \in \Theta \xrightarrow[\varphi]{biection} \gamma \in \Gamma$
   $\theta = \varphi^{-1}(\theta) \Leftrightarrow \gamma = \varphi(\theta)$
   $\sup\limits_{\theta} L(X,\varphi(\gamma)) = \sup\limits_{\gamma} L(x,\gamma)$
   $\gamma* = \varphi(\theta*)$
   Пример
   Пусть у нас есть $Exp(\lambda)$ и есть две параметризации

   - $\lambda e^{-\lambda x} \to \frac{1}{\overline{\overline{X}}}$
   - $\frac{1}{\lambda}\exp(-\frac{x}{\lambda}) \to \overline{X}$

Теорема Ассимптотическая нормальность ОМП
Пусть наша модель регулярная, так же пусть:
$|\frac{\partial^3 \ln f(x,\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}| \leq M$
θ_* - ОПМ для θ
Уравнение $\nabla \ln L(X,\theta) = 0$ имеет еддинственное решение. Тогда:

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

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

В прошлый раз мы ввели функцию
$V(X,\theta) = \dfrac{\partial \ln L(X,\theta)}{\partial\theta}$
θ_0 - реальный параметр
Давайте напишемя ряд Тейлора
$V(X,\theta) = V(X,\theta_0) + V'_\theta(X,\theta)(\theta-\theta_0) + V''_\theta(X,\tilde{\theta})\dfrac{(\theta-\theta_0)^2}{2}$, $\tilde{\theta}$ между θ_0 и θ
Выполним подстановку $\theta = \theta_*$
$0 = V(X,\theta_0) + V'_\theta(X,\theta_0)(\theta_*-\theta_0) + V''_\theta(X,\tilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2}$
$V'_\theta(X,\theta_0)(\theta_*-\theta_0) = -V(X,\theta_0) - V'(X,\tilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2}$
$\sqrt{n}V'_\theta(X,\theta_0)(\theta_*-\theta_0) = -\sqrt{n}V(X,\theta_0) - \sqrt{n}V''_\theta(X,\tilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2}$
$A_n := -\sqrt{n}V(X,\theta_0)$
По ЦПТ:
$A_n \to N(0, i(\theta_0))$
$\sqrt{n}V''_\theta(X,\tilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2} = n^{\frac{3}{2}} - \dfrac{V''_\theta(X,\tilde{\theta})}{n}\dfrac{(\theta_*-\theta_0)^2}{2}$
$\dfrac{V''_\theta(X,\tilde{\theta})}{n}$ - огр по ЗБЧ
$\sqrt{n}V''_\theta(X,\tilde{\theta})\dfrac{(\theta_*-\theta_0)^2}{2} \to N(0, i(\theta_0))$
$V'(X,\theta_0) = n\dfrac{V'(X,\theta_0)}{n} \xrightarrow{\text{ЗБЧ}} -i(\theta)$
$\operatorname{Var}\widehat{\theta} \geq \dfrac{[\tau'(\theta)]^2}{ni(\theta)}$
Рассмотрим показатель:
$\dfrac{[\tau'(\theta)]^2}{ni(\theta)\cdot\operatorname{Var}\widehat{\tau(\theta)}}$ - Эффективность
Ассимптотическая Эффективность: Пусть $\sqrt{n}(\widehat{\theta}-\theta_0) \to N(0, \dfrac{\sigma^2}{n})$
$\dfrac{1}{i(\theta)\sigma^2}$ - показатель состоятельной эффективности

11 Экспоненциальное семество 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: $N(), \Gamma(), Pois(), Bin, NB$
$\ln p(x,\theta) = A(\theta)B(x) + C(\theta) + D(x)$

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

12 Байесовская постановка

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

13 Минимаксные оценки

$m(\widehat{\theta}) = \underset{\theta}{sup} R(\widehat{\theta}, \theta)$

$\widehat{\theta_{WC}} = argmin m(\widehat{\theta})$ - минимаксная оценка

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

Утверждение

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

Рассмотрим пример:

$Bern(p)$: prior: B(a, b)

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

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

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

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

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

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

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

14 Интервальное оценивание

Определение. Доверительный интервал

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

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

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

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

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

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) $-\ln F_\theta(X_k) \sim Exp(1)$

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

в) $-\sum \ln F_\theta(X_k) \sim \Gamma(n,1)$
Доверительные интервалы нормального закона. Теорема Фишера
Лемма о независимотси линейной и квадратичной статистик
$X_1, \dots, X_n \sim N(\mu, \sigma^2)$
$T = AX, X = (X_1, \dots, X_n)^T A \in M_{m\times n}(\mathbb{R})$
$Q = X^T BX, B \in M_n(\mathbb{R}), B = B^T$
Тогда T, Q - независимы
$AB = 0$
Доказательство
$\Lambda = U^T BU$
$\Lambda = diag(\lambda_1, \dots, \lambda_m, 0, 0)$
λ_k - собственное число не 0
$U = (u_1, \dots, u_n)$ - собственные векторы ортонормированного базиса $\Leftrightarrow B = U\Lambda U^T = \sum\limits_{j=1}^{m} \lambda_j u_j u j^T \Rightarrow Q = \sum\limits_{j=1}^{M} \lambda_j (X^T U_j)(U_j^T X) = \sum\limits_j \lambda_j (U_j^T X)^2$
$A(\sum\limits_{j=1}^{m} \lambda_j u_j u_j^T) = 0$
$\sum\limits_{j=1}^{m} \lambda_j A U_j u_j^T = 0$
Зафиксируем $1 \le k \le m$ домножим справа на u_k
$Au_k = 0 \Rightarrow \forall i A[i,*]u_k = 0$
Нам надо доказать, что $\forall i, k A[i,*]X$ и $u_k^T X$ - нез
$\mathrm{Cov}(A[i,*]X, u_k^T X) = \mathrm{Cov}(A[0,*]X, X^T u_k) = A[i,*] \operatorname{Var} X u_k = \sigma^2 A[i,*]u_k = 0$
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
$X_1, \dots, X_n \sim N(0,1)$
$\sum\limits_{k=1}^{n} X_k^2 \sim \Xi^2(n)$ (распределение)
хи-кквадрат с n степенями свободы
Лемма о распределении квадратичной статистики
$X_1, \dots, X_n \sim N(0,1)$
$Q = X^T BX$
$B = B^2$
Тогда $Q \sim \Xi^2(r), r = rank(B = tr(B))$
$Q = \sum\limits_{k=1}^{n} (u_k^T X)^2 \sim \Xi^2(r)$
$u_k^T \sim N(u_k^T EX, u_k^T I_n u_k) = N(0,1)$

$\text{Cov}(u_k^T X, u_j^T X) = 0$
$B = U\Lambda U^T$
$rankB = rank\Lambda = tr\Lambda$
$trB = tr(U\Lambda U^T)$
Заметим что $B_{j,j} = \lambda_j u_j u_j^T = \lambda_j$
Теорема Фишера
$X_1, \ldots, X_n \sim N(\mu, \sigma^2) \Rightarrow$

1. $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
2. $\frac{nS_*^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim X^2(n-1)$
3. $S^2, \overline{X}$ - нез

$Y_j = \frac{X_j - M}{\sigma}$
$\overline{X}\frac{1}{\sigma}(\overline{X} - M)$
$S_*^2(Y) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \overline{Y})^2 = \frac{S_*^2(X)}{\sigma^2}$
$\overline{Y} = \frac{\sum Y_j}{n} = \frac{(\frac{1}{n}, \ldots, \frac{1}{n})}{b}(Y_1, \ldots, Y_n)^T = bY$
$nS_*^2(Y) = (Y - BY)^T(Y - BY) = Y^T(I-B)^T(I-B)Y \sum \Xi^2(tr(I-B))$
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, $\mu = ?$
  Рассмотрим два варианта: $\frac{X_1 - \mu}{\sigma} \sim N(0,1)$
  $\sqrt{n}\frac{\overline{X} - \mu}{\sigma} \sim N(0,1)$
  Доверительный интервал уровня $1 - \alpha$ $[\overline{X} - \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \overline{X} + \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}]$

- μ - известно, $\sigma^2 = ?$
  $-q \leq \sqrt{n}\frac{(\overline{X} - \mu)}{\sigma} \leq q$
  $-q\sigma \leq \sqrt{n}(\overline{X} - \mu) \leq q\sigma$
  $\frac{\sqrt{n}(\overline{X} - \mu)}{q} \leq \sigma$

$$-\frac{\sqrt{n}(\overline{X}-\mu)}{q} \le \sigma$$

Рассмотрим следующую статистику:

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

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

Доверительный интервал:

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

Давайте теперь рассмотрим задачу построения доверительного интревала

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

Воспользуемся теоремой Фишера:

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

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

Определение. Распреление Стьюдента

$X_0, X_1, \dots, X_n$ - нез, $N(0,1)$

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

n - степени свободы (deg of freedom)

Давайте выведем статистику:

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

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

Доверительный интервал:

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

Определение. Распределение Фишера

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

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

Они независимы

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

15 Ассимптотические доверительные интервалы

Раньше мы говорили $P(\theta \in (l_n, r_n)) \geq 1 - \alpha$
Таперь же мы будем говорить $\lim\limits_{n\to\infty} P(\theta \in (l_n, r_n)) \geq 1 - \alpha$

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

1. ЦПТ и ее следствия
   $\sqrt{n}\dfrac{\overline{X} - \mu}{\sigma} \to N(0,1)$
   $\sqrt{n}\dfrac{\overline{X} - \mu}{S} \to N(0,1)$
   Доверительный интервал: $\overline{X} \pm \dfrac{q_{1-\frac{\alpha}{2}} S}{\sqrt{n}}$
   $\sqrt{n}\dfrac{S_*^2 - \sigma^2}{\sqrt{\widehat{\beta_4} - S_*^4}} \to N(0,1)$
   Доверительный интервал: $S_*^2 \pm \dfrac{q_{1-\frac{\alpha}{2}}\sqrt{\widehat{\beta_4} - S_*^4}}{\sqrt{n}}$
2. Теорема об ассимтотике среднего члена вариационного ряда
   $\sqrt{n}\dfrac{X_{(\lfloor np \rfloor)} - q_p}{\sqrt{p(1-p)}} f(q_p) \to N(0,1)$
   Доверительный интервал для медианы: $p = \dfrac{1}{2}$
   $\sqrt{n} f(q_p) \dfrac{X_{(\lfloor \frac{n}{2} \rfloor)} - q_p}{\frac{1}{2}}$ (зачастую f это константа)
   Доверительный интервал: $X_{(\lfloor \frac{n}{2})\rfloor} \pm \dfrac{q_{1-\frac{\alpha}{2}}}{\sqrt{n} \cdot const}$
3. Доверительный интервалы из ассимтотической нормальности оценок максимального правдоподобия
4. Теорема об ассимтотике крайнего члена вариационного ряда
   $n(1 - F(X_{m+1-s})) \to \Gamma(s, 1)$
   Пример:
   $U[0, \theta]$
   $F(x) = \begin{cases} 0, x < 0 \\ \dfrac{x}{\theta}, 0 \leq x \leq \theta \\ 1, x > \theta \end{cases}$
   $\dfrac{nX_{(r)}}{\theta} \to \Gamma(r, 1)$

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

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 континентальное 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.
$X_1, \ldots, X_n$ - выборка в широком смысле.
$(X_1, \ldots, X_n) \sim F$
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)
$\delta(X, \alpha, H_0, H_1) = \begin{cases} accept H_0 \\ reject H_0 (w.\ respect \text{ to } H_1) \end{cases}$ 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 $\to$)
$P(T(X) \in T_0(\alpha)|H_0) = 1 - \alpha$
if $T(x) \in T_1(\alpha):$ reject H_0
else: acceptH_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: $T_0(\alpha) = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}], T_1(\alpha) = \overline{T_0(\alpha)}$

$p_l = P(U \leq T(x)|H_0)$
$p_r = P(U > T(x)|H_0)$
$p = 2min(p_l, p_r)$
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.
$X_1, \ldots, X_n \sim F_\theta$
$T(X, \theta) \to U \sim G$
$P(q_{\frac{\alpha}{2}} \leq T(X, \theta) \leq q_{1-\frac{\alpha}{2}}) = 1 - \alpha$
$H_0 : \theta = \theta_0$
$P(T(X, \theta_0) \in T_0(\alpha)|\theta = \theta_0) = 1 - \alpha$
$H_1 = \theta \neq \theta_0, \theta > \theta_0, \theta < \theta_0$
Пример:
1) $X_1, \ldots, X_n \sim F, \mu = EX_1, \exists \operatorname{Var} X$
$H_0 : \mu = \mu_0$
$T(X) = \sqrt{n}\dfrac{\overline{X} - \mu_0}{S} \to N(0, 1)$ если $\mu = \mu_0$
При $H_1 : \mu \neq \mu_0$ у нас двухсторонняя критическая область
При $H_1 : \mu > \mu_0$ у нас правостороння критическая область
При $H_1 : \mu < \mu_0$ у нас левосторонняя критическая область
Дальше 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)

$P(\sqrt{n}\frac{\overline{X}-\mu_0}{S} \in T_0(\alpha)|\mu \neq \mu_0) = P(\sqrt{n}\frac{\overline{X}-\mu}{S} + \frac{\mu-\mu_0}{S}\sqrt{n}|\mu \neq \mu_0)$
Это будет стремиться либо к $\Phi(-\infty)$ либо $1 - \Phi(+\infty)$

18 Критерий Колмогорова

$X_1, \dots, X_n \sim F$
$H_0 : F = F_0$ (F_0 - непр)
$H_1 : F \neq F_0$
Идея основана на теореме Колмогорова (было в начале семестра)
$D_n = \sqrt{n} \underset{x \in \mathbb{R}}{sum} |F_n(x) - F_0(x)|$, F_n - 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 приблеженные формулы для D_n
3) $H_0 : F = F(\theta), H_1 : \neg H_0 \Rightarrow D_n = \sqrt{n} \underset{x}{sup} |F_n(x) - F_0(x, \theta)|$
$\theta \to \widehat{\theta}$
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
$D_{n,m} = \sqrt{n} \underset{x}{sup} |F_n(x) - F_m(x)|$
$T_1(\alpha) = (q_{1-\alpha}, +\infty)$

20 Критерий типа хи-квадрат

20.1 Критерий согласия Пирсона

$X_1, \dots, X_n \sim F(x)$ - непрерывная
Давайте ддискритезируем данные
$\Delta_1 : \nu_1$ - количесво элементов выборки попадающих Δ_1
$\dots$
Δ_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
$H_0 : p = p_0$
$H_1 : p \neq p_0$
$n = \sum\limits_{k=1}^{N} \nu_k$
$\Xi_N^2 = \sum\limits_{k=1}^{N} \frac{(\nu_k - np_{0k})^2}{np_{0k}}$
Теорема
$\Xi_N^2 \xrightarrow[n \to \infty]{} \Xi^2(N-1)$ при условии H_0
Доказательство
$N = 2 :$
$\frac{(\nu_1 - np_{01})^2}{mp_{01}} + \frac{(\nu_2 - np_{02})^2}{np_{02}} = \frac{(\nu_1 - np_{01})^2}{np_{01}} + \frac{(n - \nu_1 - n(1-p_{01}))^2}{n(1-p_{01})} = \frac{(\nu_1 - np_{01})^2}{n}(\frac{1}{p_{01}} + \frac{1}{1 - P_{01}}) = \frac{(\nu_1 - np_{01})^2}{np_{01}(1-p_{01})}$
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 притерий хи-квадрат у нас правосторонняя критическая область (потому что в сумме при H_0 будет маленькая разность в квадртае 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)
$\nu = (315, 101, 108, 32)$
$p_0 = (\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16})$
$\Xi_4^2 = 0.47$
$df = 3$
$p\_value = 1 - (df(0.47)) = 0.97$
Замечание
$n \geq 50, \nu_j \geq 5$ **Замечание**
Оптимальные способы дискритизации сущестуют
Замечание
Для проверки согласованности с нормальным законом следует использовать спец тесты (Шапиро, коэффициент эксцесса)
$H_0 : p = p_0(\theta), \theta \in \Theta \subset \mathbb{R}^d, d < N - 1$
$\Xi_N^2 = \sum\limits_{k=1}^{N} \frac{(\nu_k - np_{0k}(\theta))^2}{np_{0k}(\theta)}|_{\theta=\widehat{\theta}}$
Теорема
$p_0(\theta) > 0 \forall \theta$
$\frac{\partial p_0}{\partial \theta}, \frac{\partial^2 p_0}{(\partial \theta)^2}$ - непрерывная
Тогда $\Xi_N^2 \to X^2(N - 1 - d)$

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

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
$n_i = \sum\limits_j \nu_{ij} = \nu_{i*}$ - объем i-ой выборки
$n = n_1 + \dots + n_k$
Пусть $p^{(1)}, \dots, p^{(k)}$ - известны
$\Xi^2_{n_1} = {}^N_{j=1} \frac{(\nu_{ij} - n_i p_j^{(i)})^2}{n_i p_j^(i)}, df = N - 1$
$\Xi^2_{n_1, n_2, \dots, n_k} = \sum\limits_{i=1}^K \Xi^2_{n_i}, df = k(N-1)$
Рассмотрим $p^{(1)}, \dots, p^{(k)}$ - не известны
$df = k(N-1) - (N-1) = (N-1)(k-1)$
$\widehat{p_j} = \frac{\nu_{1j} + \dots + \nu_{kj}}{n} = \frac{\nu_{*j}}{n}$
$L(\dots) = p_1^{(\nu_{*1})} \dots p_k^{(\nu_{*k})}, n = \sum\limits_j \nu_{*j}$
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
Проссумируем по каждому столбцу и по каждой строчке
Пусть $p_{ij} = P(X = i, Y = j)$
$p_{xi} = P(X = i)$
$p_{yj} = P(Y = j)$
$H_0 : p_{ij} = p_{xi} p_{yj} \forall i, j$
$H_1 : \neg H_0$
$\Xi^2 = \sum\limits_{1 \leq i \leq N, 1 \leq j \leq M} \frac{(\nu_{ij} - p_{ij} n)^2}{n p_{ij}}, df = MN - 1$
$df = MN - 1 - (N-1) - (M-1) = MN - N - M + 1 = N(M-1) - (M-1) = (M-1)(N-1)$
$\widehat{p_{Xi}} = \frac{\nu_{i*}}{n}$

$\widehat{p_{Yj}} = \frac{\nu_{j*}}{n}$
TODO пример в коэффициентом Пирсона

21 Критерий квантилей

$H_0 :$
$F(q_1) = \alpha_1$
$F(q_2) = \alpha_2$
$\dots$
$F(q_N) = \alpha_N$
где $\alpha_0 = 0 < \alpha_1 < \dots < \alpha_N < 1 = \alpha_{N+1}$
$H_1 : \neg H_0$
$q_0 = \inf \operatorname{supp} P < q_1 < \dots < q_N < \sup \operatorname{supp} P = q_{N+1}$
$\Delta_1 = [q_0, q_1)$
$\dots$
$\Delta_{N+1} = [q_N, q_{N+1})$
$P(\Delta_1) = \alpha_1 - \alpha_0$
$\dots$
$P(\Delta_N) = \alpha_{N+1} - \alpha_N$
Рассмотрим особый случай $H_0 : F(q) = \frac{1}{2}$

22 Критерий знаков

$(X_1, Y_1)^T, \dots, (X_n, , Y_n)^T$
Мы хотим проверить что:
а) выборки независимы
б) распределения одинаковы
$F(x, y) = F_1(x) \cdot F_1(y)$
$U = X - Y \Rightarrow \operatorname{med} U = 0$
ν_1 - количество элементов новой выборки $>$ med
$Z_n = \frac{2}{\sqrt{n}}(\nu_1 - \frac{n}{2}) \to N(0, 1)$
На предыдущей лекции мы обсудили:
$Z_n = \sqrt{n}\rho_n, \rho_n$ - коэффициент корреляции Пирсона
Теорема
$(X_1, Y_1)^T, \dots, (X_n, , Y_n)^T \sim N(\dots, \dots) \Rightarrow \frac{\sqrt{n-2}\rho_n}{\sqrt{1-\rho_n^2}} \sim T(n-2)$
$H_0 : \rho = 0$
$H_1 : (\rho \neq 0, \rho > 0, \rho < 0)$

23 Ранговые критерии

Определение. Ранг
$X_1, \dots, X_n$ - выборка
$r(X_k)$ - номер X_k в вариационном ряде
$X_1, \dots, X_n$
$Y_1, \dots, Y_m$
Это две независимных выборки, давайте объединим их в одну
Рассмотрим $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ - Статистика Вилкоксона
R_i - ранг X_i, в объединенной выборке
$T = R_1 + \dots, R_n$
$Z_{rs} = \mathbb{1}(X_r < Y_s)$
$U = \sum\limits_{r=1}^{n}\sum\limits_{s=1}^{m} Z_{rs}$ - Мант-Уитни
$T + U = mn + \frac{n(n+1)}{2}$
Хотим проверить, что распределение X совпадает с Y
$EU = mnE\mathbb{1}(X < Y) = mnP(X < Y) = \frac{1}{2}$ - при условии H_0
$H_0 : P(X < Y) = \frac{1}{2}$
$U \sim N(\frac{mn}{2}, \frac{nm(m+n+1)}{12})$
$(X_1, Y_1)^T, \dots, (X_n, , Y_n)^T$
Хотим проверить независимость
R_i - ранг X_i (в своей выборке)
S_i - ранг Y_i (в своей выборке)
ρ - выборочный коэффициент корреляции между R_i и S_i - коэффициент корреляции Спирмена
$\rho = \frac{12}{n(n^2-1)} \sum (R_i - \frac{n+1}{2})(S_i - \frac{n+1}{2})$
$H_0 : \rho = 0, \sqrt{n}\rho \to N(0, 1)$
$H_1 : \rho \neq 0, \rho > 0, \rho < 0$
Коэффициент корреляции Кенделла
$\tau = \frac{2}{n(n-1)} \sum\limits_{i=1}^{n-1} {}_{j=i+1}^{n} \operatorname{sign}(T_j - T_i)$
H_0 верно $\Rightarrow E\tau = 0, \operatorname{Var} \tau = \frac{2(2n+5)}{9n(n-1)}$
$\tau \approx N(0, \frac{4}{9n})$
$H_0 : \tau = 0$
$H_1 :> 0, < 0, \neq 0$

24 Критерий инверсии

$X_{(1)} \leq \dots \leq X_{(n)}$ Крайние ситуации: выборка отсортированна, то есть трудно поверить, что у нас все случайно
ν_i - количество инверсий для элемента X_i
$T = \nu_1 + \dots + \nu_n$
$ET = \dfrac{n(n-1)}{4}$
$\operatorname{Var} T = \dfrac{n(n-1)(2n+5)}{72}$
Статистика T - ассимптотически нормальная

25 Линейные статистические модели

$Y = Xb + \varepsilon$
$X \in M_{n\times m}(\mathbb{R})$ - матрица переменных
x_{ij} - количественная переменная
$Y \in \mathbb{R}^n$ - наблюдения зависимой переменной
$b \in \mathbb{R}^m$ - неизвестный вектор коэффициентов

$varepsilon \in \mathbb{R}^n$ - случайная ошибка
1) $E\varepsilon = 0$
2) $\operatorname{Var} \varepsilon_i = \sigma^2$ - гомоскедотичность
3) $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0$
Наша цель оценить b и σ^2
Определение. Ошибка наименьших квадратов
$\widehat{b} = \operatorname{argmin} S^2(b)$
$S^2(b) = (Xb - Y)^T(Xb - Y)$
$A = X^TX \in M(m \times n)$
$\dfrac{1}{n}X^TX$
$rank(A) = m$
Теорема
$\widehat{b} = A^{-1}X^TY$
Доказательство
$S^2(\widehat{b} + \delta) = S^2(\widehat{b}) + \delta^T A\delta$
$t = Tb, T \in M_{k\times m}, k \leq m, rankT = k$
$\widehat{t} = T\widehat{b}$
Теорема. Гаусс-Марков
1) $\widehat{t}$ - несмещенная оценка t
2) $\widehat{t}$ - оптимальная оценка в классе линейных несмещенных оценок
Доказательство
$E\widehat{t} = ET\widehat{b} = ETA^{-1}X^TY = TA^{-1}X^TEY = TA^{-1}X^TXb = Tb = t$
$\operatorname{Var}(\widehat{t}) = \operatorname{Var} TA^{-1}X^TY = TA^{-1}X^T \operatorname{Var} Y XA^{-1}T^T = \sigma^2 TA^{-1}T^T$
$\operatorname{Var} \widehat{t} = \sigma^2 TA^{-1}T^T$

$\operatorname{Var}\widehat{b} = \sigma^2 A^{-1}$
Пусть LY - несмещенная оценка для t, т.е.
$ELY = t = Tb \Rightarrow \operatorname{Var} LY = \sigma^2 LL^T$
Заметим, что $LL^T = (TA^{-1}X^T)(TA^{-1}X^T)^T + (L - TA^{-1}X^T)(L - TA^{-1}X^T)^T$
$MSE\widehat{t} = tr\operatorname{Var}\widehat{t}$
Тогда $L = TA^{-1}X^T$

$ES^2(b) = E(Xb-Y)^T(Xb-Y) = E\varepsilon^T\varepsilon = n\sigma^2$
$E(\widehat{b}-b)A(\widehat{b}-b) = \sum\limits_{i,j} a_{ij}E(\widehat{b}_i-b_i)(\widehat{b}_j-b_j) = \sum\limits_{i,j} a_{ij}\operatorname{Cov}(\widehat{b}_i,\widehat{b}_j) = \sigma^2\sum\limits_{i,j} a_{ij}a^{-1}_{ij} = \sigma^2\sum\limits_i {}_j a_{ij}a^{-1}_{ji} = \sigma^2{}^m_{i=1}1 = \sigma^2 m$

$S^2(b) = S^2(\widehat{b}) + (\widehat{b}-b)^T A(\widehat{b}-b) \Rightarrow n\sigma^2 = ES^2(\widehat{b}) + m\sigma^2 \Rightarrow \widehat{\sigma}^2 = \dfrac{S^2(\widehat{b})}{n-m}$ - несмещенная оценка для σ^2

26 Условные оценки наименьших квадратов

Определение
$Tb = t_0, T \in M_{k\times m}, rankT = k$
$\widehat{b}_{T,t_0} = \underset{Tb=t_0}{\operatorname{argmin}} S^2(b)$ - условная оценка наименьших квадратов
Теорема
$\widehat{b}_{T,t_0} = \widehat{b} - A^{-1}T^TD^{-1}(T\widehat{b}-t_0), D = TA^{-1}T^T$
Можно показать: $S^2(b) = S^2(\widehat{b}_{T,t_0}) + (\widehat{b}_{T,t_0}-b)^TA(\widehat{b}_{T,t_0}-b)$
Замечание: $S^2(b) = S^2(\widehat{b}) + (\widehat{b}-b)^TA(\widehat{b}-b)$
$b = \widehat{b}_{T,t_0} : S^2(\widehat{b}_{T,t_0}) = S^2(\widehat{b}) + (\widehat{b}-\widehat{b}_{T,t_0})^TA(\widehat{b}-\widehat{b}_{T,t_0})$
$(T\widehat{b}-t_0)^TD^{-1}(T\widehat{b}-t_0) = S^2(\widehat{b}_{T,t_0}) - S^2(\widehat{b})$
"Обычное предположение": $\varepsilon \sim N(0,\sigma^2E_n)$
$Y = Xb + \varepsilon \sim N(Xb, \sigma^2E_n)$
$L(b,\sigma^2,Y) = \dfrac{1}{(\sqrt{2\pi\sigma^2})^n}\exp(-\dfrac{1}{2\sigma^2}(Xb-Y)^T(Xb-Y))$
$\sup L(b,\sigma^2,Y) \Leftrightarrow \inf S^2(b)$
$\widehat{b}$ - эффективна в классе несмещенных оценок $\Leftarrow \widehat{b}$ - о.м.п
Теорема. Основная теорема о нормальной регрессии
1) $S^2(\widehat{b}), \widehat{b}$ - нез.

27 Последовательный анализ Вальда

$H_0 : F = F_0$
$H_1 : F = F_1$ f_0, f_1 - соответствующие плотности $A_0 < 1 < A_1$

$n = 1$

$\alpha, \beta, a_0 = \ln A_0, a_1 = \ln A_1$
$Z_i = \ln \dfrac{f_1(X_i)}{f_0(X_i)}$
$\sum_{i=1}^{n} Z_i = \ln \dfrac{L_{1n}}{L_{0n}}$
Теорема
Пусть ν - количество итераций $\Rightarrow P(@n \geq n|H_0) \to 0, P(\nu \geq n|H_1 \to 0), n \to \infty$
Доказательство
$r - fixed$
$\eta_1 = Z_1 + \dots + Z_r$
$\eta_1 = Z_{r+1} + \dots + Z_{2r}$
$\dots \ \{\nu \geq rk\} \Leftrightarrow \{a_0 < \eta_1 + \dots + @n_j < a_1\}_{i \leq j \leq k} = \{a_0 < \eta_1 < a_1, a_0 < \eta_1 + \eta_2 < a_1, \dots, a_0 < \eta_1 + \dots + \eta_k < a_1\} \subset \{|\eta_j| < b = a_i - a_0\}_{1 \leq j \leq k} \Rightarrow P(\nu \geq rk|H_s) \leq P_{1 \leq j \leq k}(|\eta_j| < b|H_s) = P_{1 \leq j \leq k}(\nu_j^2 < b^2|H_s) = P^k(\eta_1^2 < b^2|H_s) = p_s^k$
$E_s \eta_1^2 \geq \operatorname{Var}_s \eta^1 = r \operatorname{Var}_s Z_1 > b^2, r > \max\left(\dfrac{b^2}{\operatorname{Var}(Z_1|H_0)}, \dfrac{b^2}{\operatorname{Var} Z_1|H_1}\right)_{\substack{E_s(\dots)=E(\dots|H_s) \\ \operatorname{Var}_s(\dots)=\operatorname{Var}(\dots|H_s)}} \Rightarrow$
$p + s < 1$
$P(\nu > n|H_s) \leq P(\nu \geq rk|H_s) \leq p_s^k \to 0, k = k(n), n \to \infty$

Теорема
Пусть (α, β) - вероятности $\Rightarrow A_0 \geq A_0' = \dfrac{\beta}{1-\alpha}$ и $A_1 \leq A_1' = \dfrac{1-\beta}{\alpha}$
if A_0' и $A_1 = A_1'$ then α', β' - вероятности ошибок
$\alpha' \leq \dfrac{\alpha}{1-\beta}, \beta' \leq \dfrac{\beta}{1-\alpha}$
$\alpha' + \beta' \leq \alpha + \beta \ 1 = \sum_{n=1}^{\infty} P(\nu = n|H_0) = \sum_{n=1}^{\infty} P(\text{accept at the step n} \mid H_0) + \sum_{n=1}^{\infty} P(\text{reject at the step n} \mid H_0) = \alpha = \sum n = 1^{\infty} P(\text{reject at the step n} \mid H_0) \leq \dfrac{1}{A_1} \sum_{n=1}^{\infty} P(\text{reject at the step n} \mid H_1)$
$L_{n1} \geq L_{n0} \cdot A_1$
$L_{n1 \geq L_{n0} \cdot A_1}$
Аналогичным образом мы можем получить неравенство:
$\beta \leq A_0(1-\alpha)$
$\alpha \leq \dfrac{1}{A_1}(1-\beta)$
Пусть $A_0' = \dfrac{\beta}{1-\alpha}$ и $A_1' = \dfrac{1-\beta}{\alpha}$ и ошибки $(\alpha', \beta') \Rightarrow \dfrac{\beta}{1-\alpha} \geq \dfrac{\beta'}{1-\alpha'}$
$\dfrac{1-\beta}{\alpha} \leq \dfrac{1-\beta'}{\alpha'} \Rightarrow \alpha' \leq \dfrac{\alpha(1-\beta')}{1-\beta} \leq \dfrac{\alpha}{1-\beta}$

$\beta(1-\alpha) \leq \beta(1-\alpha')$
$(1-\beta)\alpha' \leq (1-\beta')\alpha$
Сложим последние два неравества и получим требуемое
Давайте вспомним тождество Вальда:
(X_i) - независимые случайные величины с математическим ожиданием a
ν - целочисленная случайная величина не зависящая от X_i
$E\nu = b$
$S_\nu = X_1 + \cdots + X_\nu$
$\Rightarrow ES_\nu = ab = E(ES_\nu|\nu) = E\nu a = aE\nu = ab$
$a_0 = \ln\dfrac{\beta}{1-\alpha}, a_1 = \ln\dfrac{1-\beta}{\alpha}$
$E_s\nu \cdot E_s Z_i = E_s S_\nu \approx a_0(1-\alpha) + a_1(1-\beta)$
$E_1\nu \cdot E_1 Z_i \approx \beta a_0 + a_1(1-\alpha)$