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 - $X_1, X_2, \ldots, X_n$ - $i.d.d.$, F - функция распределения (теоретическая функция).
$X_1, \ldots, X_n \sim F$ (F мы не знаем априори)
$x_1, \ldots, x_n$ - реализация выборки
Цель: оценить из реализации $x_1, \ldots, x_n$ теор F.
Эмпирическая фукнция распределения:
$\mu_n(t) = \sum\limits_{i=1}^{n} \mathbb{1}(X_i \leq t)$
$F_n(t) = \dfrac{\mu_n(t)}{n}$ - эмпирическая функция распределения.

Замечание: Описывает эмпирическое распределениею.

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

$\mathbb{1}(\mathbb{X_i} \leq \mathbb{t}) \sim Bern(F(t))$

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

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

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)

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

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.
$\widehat{\beta_k} = \overline{(X-\overline{X})^k} = \dfrac{1}{n}\sum_{j=1}^{n}(X_j-\overline{X})^k$ - 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.
$S_* = \overline{X^2} - (\overline{X}^2)$
$\widehat{\beta_k} = Poly(\widehat{\alpha_k}, \dots, \widehat{\alpha_1})$
$\widehat{\alpha_1}, \dots, \widehat{\alpha_k}$ - 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
$(\xi_n-\mu)\sqrt{n}\cdot\dfrac{1}{\sqrt{n}} \xrightarrow{d} 0$ (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.
$$P(X_{(k)} \leq t) = P(\mu_n(t) \geq k) = \sum_{j=k}^{n} C_n^j F^j(t)(1 - F(t))^{n-j}$$
$$B(z, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z t^{a-1}(1-t)^{b-1} dt = B(F(t), k, n-k+1)$$
$0 \leq z \leq 1$
Пусть $p()$ - теоретическая плотность, то есть $p = F'$
$(P(X_(k) \leq t))'_t = \dfrac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \cdot F^{k-1}(t)(1-F(t))^{n-k} \cdot p(t)$ - 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: $\frac{K(n)}{n} \to const \in (0,1)$
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:
$\sqrt{n} \cdot f(q_\alpha) \frac{X_{(\lfloor n\alpha \rfloor)} - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{d} N(0,1)$
Идея доказательства
Пусть $\lfloor n\alpha \rfloor = k$
Мы умеем писать плотность для $X_{(k)}$
Затем у нас идет преобразование:
$g(x) = \sqrt{n} p(q_\alpha) \frac{x - q_\alpha}{\sqrt{\alpha(1-\alpha)}} \rightsquigarrow p_{g(X_{(k)})}(t) = p_{X_{(k)}}(g^{-1}(t)) |g^{-1}(t)'_t|$ (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.
$Couchy(\mu, 1)$ $\alpha = \frac{1}{2}$
$2f(\mu)\sqrt{n}(X_{(\lfloor \frac{n}{2} \rfloor)}) = \frac{1}{\pi} \cdot \frac{1}{1 + (\mu - \mu)^2} \cdot 2\sqrt{n}(X_{(\lfloor \frac{n}{2} \rfloor)} - \mu)$
Теорема об 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
$b.as(\widehat{\theta}) \overset{def}{=} E\widehat{\theta} - \theta$ - 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 $g_1, \dots, g_d$
$\exists Eg_1(X_1) = m_1(\theta_1, \dots, \theta_d)$
$\exists Eg_2(X_2) = m_2(\theta_1, \dots, \theta_d)$
$\dots$
$\exists Eg_d(X_d) = m_d(\theta_1, \dots, \theta_d)$
$$\begin{cases} \overline{g_1(X)} = m_1(\widehat{\theta_1}, \dots, \widehat{\theta_d}) \\ \dots \\ \overline{g_d(X)} = m_d(\widehat{\theta_1}, \dots, \widehat{\theta_d}) \end{cases}$$
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
$U[\theta_1, \theta_2]$, $\theta_1 < \theta_2$
$g_1(x) = x, g_2(x) = x^2$
$Eg_1(X_1) = EX_1 = \dfrac{\theta_1 + \theta_2}{2}$
$Eg_2(X_1) = EX_1 = \dfrac{(\theta_2 - \theta_1)^2}{12} + \dfrac{(\theta_1 + \theta_2)^2}{4} = \dfrac{\theta_2^2 - 2\theta_1\theta_2 + \theta_1^2}{12} + \dfrac{\theta_1^2 + 2\theta_1\theta_2 + \theta_2^2}{4} = \dfrac{\theta_1^2 + \theta_1\theta_2 + \theta_2^2}{3}$
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 $(\overline{g_1(X)}, \ldots, \overline{g_d(X)})$ - состоятельаня оценка
2. если $(\overline{g_1(X)}, \ldots, \overline{g_d(X)})$ - ассимптотически нормальные и $g_1, \ldots, g_d$ - 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 $X_1, \ldots, X_n \sim p(x|\theta)$
$L(x|\theta) = \prod p(x_i|\theta)$ - 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. $N(\theta_1, \theta_2)$
   $L(x,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp\,(-\frac{(x_i-\theta_1)^2}{2\theta_2})$
   $\ln L(x,\theta) = \sum_{i=1}^{n} [-\frac{1}{2}\ln 2\pi - \frac{1}{2}\theta_2 - \frac{(x_i-\theta_1)^2}{2\theta_2}]$
   $\frac{\partial \ln L(x,\theta_1)}{\theta_1} = \sum_{i=1}^{n} \frac{2(x_i-\theta_1)}{2\theta_2} = \sum_{i=1}^{n} \frac{x_i-\theta_1}{\theta_2}$
   $\frac{\partial \ln L(x,\theta)}{\partial\theta_2} = \sum_{i=1}^{n} [-\frac{1}{2\theta_2} + \frac{(x-i-\theta_1)^2}{2\theta_2^2}]$
   $\sum_{i=1}^{n} \frac{(x_i-\widehat{\theta_1})}{\widehat{\theta_2}} = 0 \Rightarrow \widehat{\theta_1} = \overline{X}$
   $\sum_{i=1}^{n} [-\frac{1}{2\widehat{\theta_2}} + \frac{(x_i-\widehat{\theta_1})^2}{2\widehat{\theta_2}^2}] = 0 \Rightarrow \widehat{\theta_2} = S_*^2$

2. $U[0,\theta] : L(X,\theta) = \frac{1}{\theta^n}\mathbb{1}(X_i \in [0,\theta], \forall 1 \leq i \leq n)$
   $X_1, \ldots, X_n : p(X_i, \theta)$
   $L(X_1, \ldots, X_n, \theta) = \prod_{i=1}^{n} p(x_i, \theta)$
   $L(X,\theta) = \theta^{\Sigma X_i}(1-\theta)^{n-\Sigma X_i}$
   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
$L(X,p) = p_1 \ldots p_m$
$\ln L(X,p) = \sum_{j=1}^{m-1} \nu_j \ln p_j + \nu_m \ln(1 - P_1 - \cdots - p_{m-1})$
$\frac{\partial \ln L \ldots}{\partial p_j} = \frac{\nu_j}{p_j} - \frac{\nu_m}{1-p_1-\cdots-p_{m-1}} = 0$
$\sum$ 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:
$i(\theta) = -(E \frac{\partial^2 \ln p(X, \theta)}{\partial \theta_i \partial \theta_j})_{1 \leq i, j \leq d}$
$I(\theta) = n i(\theta)$
$N(\theta_1, \theta_2)$

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

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

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

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

10 Состоятельность ОМП

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

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

Теорема Ассимптотическая нормальность ОМП
Пусть наша модель регулярная, так же пусть:
$|\dfrac{\partial^3 \ln f(x,\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k}| \le M$
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: $N(), \Gamma(), Pois(), Bin, NB$
$\ln p(x,\theta) = A(\theta)B(x) + C(\theta) + D(x)$

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

12 Байесовская постановка

$X_1, \dots, X_n \sim F_\theta$
$\theta \sim \pi(\theta)$ - 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:
$P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$
$P(\theta|X) = \dfrac{L(X|\theta)\pi(\theta)}{\displaystyle\int L(X|\theta)\pi(\theta)d\theta}$
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)

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

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

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

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

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

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

- Достаточность и полные статистики
- Робастность (устойчивость относительно выборосов, 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

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
$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$
$T = AX, X = (X_1, \ldots, X_n)^T A \in M_{m \times n}(\mathbb{R})$
$Q = X^T BX, B \in M_n(\mathbb{R}), B = B^T$
Тогда T, Q - независимы
$AB = 0$
Доказательство
$\Lambda = U^T BU$
$\Lambda = diag(\lambda_1, \ldots, \lambda_m, 0, 0)$
λ_k - собственное число не 0
$U = (u_1, \ldots, u_n)$ - собственные векторы ортонормированного базиса $\Leftrightarrow B = U\Lambda U^T = \sum\limits_{j=1}^{m} \lambda_j u_j uj^T \Rightarrow Q = \sum\limits_{j=1}^{M} \lambda_j (X^T U_j)(U_j^T X) = \sum\limits_{j} \lambda_j (U_j^T X)^2$
$A(\sum\limits_{j=1}^{m} \lambda_j u_j u_j^T) = 0$
$\sum\limits_{j=1}^{m} \lambda_j AU_j u_j^T = 0$
Зафиксируем $1 \leq k \leq m$ домножим справа на u_k
$Au_k = 0 \Rightarrow \forall i A[i, *]u_k = 0$
Нам надо доказать, что $\forall i, kA[i, *]X$ 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
$X_1, \ldots, X_n \sim N(0, 1)$
$\sum\limits_{k=1}^{n} X_k^2 \sim \Xi^2(n)$ (распределение)
хи-кквадрат с n степенями свободы
Лемма о распределении квадратичной статистики
$X_1, \ldots, X_n \sim N(0, 1)$
$Q = X^T BX$
$B = B^2$
Тогда $Q \sim \Xi^2(r), r = rank(B = tr(B))$
$Q = \sum\limits_{k=1}^{n} (u_k^T X)^2 \sim \Xi^2(r)$
$u_k^T \sim N(u_k^T EX, u_k^T I_n u_k) = N(0, 1)$

$\mathrm{Cov}(u_k^T X, u_j^T X) = 0$
$B = U\Lambda U^T$
$rank B = rank\Lambda = tr\Lambda$
$trB = tr(U\Lambda U^T)$
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. $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n})$
2. $\dfrac{nS_*^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2} \sim X^2(n-1)$
3. $S^2, \overline{X}$ - нез

$Y_j = \dfrac{X_j - M}{\sigma}$
$\overline{X}\dfrac{1}{\sigma}(\overline{X} - M)$
$S_*^2(Y) = \dfrac{1}{n}\sum_{j=1}^{n}(Y_j - \overline{Y})^2 = \dfrac{S_*^2(X)}{\sigma^2}$
$\overline{Y} = \dfrac{\sum Y_j}{n} = \dfrac{(\frac{1}{n}, \ldots, \frac{1}{n})}{b}(Y_1, \ldots, Y_n)^T = bY$
$nS_*^2(Y) = (Y - BY)^T(Y - BY) = Y^T(I-B)^T(I-B)Y \sum \Xi^2(tr(I-B))$
Для того чтобы доказать третье утверждение
$b(I-B) = b - b = 0$
Тогда мы пользуемся первой леммой
Таким образом теорема Фишера доказана.