

Математическая статистика

А.А. Дороговцев

Лекция 1

1. Пусть X_1, X_2, \dots, X_n – значения, полученные в результате измерения величины, которая является случайной.

Какое распределение имеет эта величина?

2. Наблюдается последовательность X_1, X_2, \dots, X_n – нужно проверить, являются ли X_1, X_2, \dots, X_n – реализацией независимых одинаково распределенных случайных величин?

3. Есть 2 гипотезы относительно типа распределения. Нужно выбрать одну из них.

4. Существует ли связь между параметрами x и y эксперимента и каков ее вид?

Предмет математической статистики – выяснение характера вероятностного эксперимента, отвечающего данной модели, а именно, – определение вида распределения наблюдаемой случайной величины, характеры вероятностной зависимости наблюдений последовательности случайных величин, различение гипотез, сделанных относительно эксперимента, определение вида функциональной зависимости между различными параметрами эксперимента.

Теория оценок

(занимаемся вопросом 1)

Выборочные характеристики. Понятие оценки.

Свойства оценок

Определение. Выборка – некоторое количество независимых наблюдений случайной величины. Обозначается X_1, \dots, X_n , n – объем выборки.

Если исходная случайная величина имеет распределение F , то говорят, что выборка сделана из распределения F .

Выборка – это

- 1 набор n независимых одинаково распределенных случайных величин;
- 2 n чисел.

Определение. Эмпирической функцией распределения, построенной по выборке (X_1, \dots, X_k) , называется

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{\{X_k \leq x\}}$$

Пример.

$n = 4$ X_1, X_2, X_3, X_4

Рис.

Какая связь между F_n и F ?

Теорема. $F_n \Rightarrow F, n \rightarrow \infty$ почти везде.

Лемма. $\forall x \in \mathbb{R}$

$$F_n(x) \xrightarrow{P} F(x), n \rightarrow \infty.$$

$$\triangleright y_k = \mathbb{I}_{\{x_k \leq x\}}$$

y_1, \dots, y_n, \dots – независимые одинаково распределенные случайные величины. Имеют всего 2 значения. Поэтому существует дисперсия. По закону больших чисел

$$\frac{y_1 + \dots + y_n}{n} \xrightarrow{P} My_1$$

$$My_1 = M\mathbb{I}_{\{x_1 \leq x\}} = P\{x_1 \leq x\} = F(x). \quad \triangleleft$$

Пусть X_1, \dots, X_n – выборка

Рис.

$N \ll n$ – значительно меньше.

Рассмотрим все конечные интервалы $[\Delta_k; \Delta_{k+1}]$, обозначим такой интервал $\Delta_k, k = \overline{1, N-1}$

$$\nu_i = \sum_{k=1}^n \mathbb{I}_{\Delta_i}(x_k) \quad \text{– сколько раз попали в интервал}$$

$$p_n(x) = \frac{\nu_i}{n|\Delta_i|}, x \in \Delta_i$$

Рис.

p_n – гистограмма.

Зная распределение, как определить параметры?

F_θ – семейство функций распределения, зависящих от параметра $\theta, \theta \in$

Θ – множество параметров.

Имеем выборку x_1, \dots, x_n . Неизвестный параметр θ заменим некоторой функцией выборки $\Phi(x_1, \dots, x_n)$.

Определение. $\Phi(x_1, \dots, x_n)$, построенная для определения θ по выборке x_1, \dots, x_n , называется оценкой θ и обозначается обычно $\hat{\theta}, \theta', \theta^x, \dots$.

Примеры.

1. x_1, \dots, x_n – выборка из распределения

$$\begin{array}{c|c} 0 & 1 \\ \hline 1-p & p \end{array} \quad p - \text{параметр, } p \in [0; 1]$$

$$p' = p(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n} = \bar{x}$$

$$p'' = \sqrt[n]{x_1 \cdot \dots \cdot x_n}$$

$$p''' = x_1.$$

Что происходит с оценкой при увеличении N ?

$$p' \xrightarrow{P} p, \quad n \rightarrow \infty$$

p'' превращается в 0 п.н., $n \rightarrow \infty$

$p''' \not\rightarrow p, n \rightarrow \infty$.

Определение. Оценка $\hat{\theta}$ называется состоятельной, если

$$\hat{\theta} \xrightarrow{P} \theta, \quad n \rightarrow \infty.$$

(Если имеет место сходимость п.н., то говорят, что оценка сильно состоятельна).

Определение. Оценка $\hat{\theta}$ называется несмещенной, если

$$M_{\theta} \hat{\theta} = \theta, \quad \forall \theta \in \Theta$$

($\hat{\theta}$ распределение с параметром θ).

Упражнение. Не существует несмещенной оценки для параметра $\frac{1}{\lambda}$ для пуассоновского распределения по одному наблюдению ($n = 1$),

\bar{x} – выборочное среднее (арифметическое).

\bar{x} – несмещенные оценки настоящего среднего.

Выборочная дисперсия: $\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$.

Упражнения. 1. Проверить, что $\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$ – смещенная оценка.

2. Найти правильный множитель.

3. Проверить, если есть старшие моменты, что она состоятельна.

Лекция 2

$M_\theta(\hat{\theta} - \theta)^2$ – среднеквадратическое отклонение несмещенной оценки от параметра.

Определение. Оценка θ_* называется оптимальной, если для каждой оценки $\hat{\theta}$ справедливо неравенство

$$M_\theta(\hat{\theta} - \theta)^2 \geq M_\theta(\theta_* - \theta)^2, \quad \theta \in \Theta$$

Считаем все оценки несмещенными.

Теорема. Если оптимальная оценка существует, то она единственная.

▷ θ_* – оптимальная несмещенная оценка.

Пусть существует другая оптимальная несмещенная оценка θ' . Рассмотрим

$$\hat{\theta} = \frac{1}{2}(\theta_* + \theta')$$

$\hat{\theta}$ – несмещенная.

$$\begin{aligned} M_\theta(\hat{\theta} - \theta)^2 &= \frac{1}{4}M_\theta(\theta_* - \theta)^2 + \\ &+ \frac{1}{2}M_\theta(\theta_* - \theta)(\theta' - \theta) + \frac{1}{4}M_\theta(\theta' - \theta)^2 \leq \\ &\leq \frac{1}{4}M_\theta(\theta_* - \theta)^2 + \frac{1}{2}\sqrt{M_\theta(\theta_* - \theta)^2}\sqrt{M_\theta(\theta' - \theta)^2} + \frac{1}{4}M_\theta(\theta' - \theta)^2 \end{aligned}$$

так как θ_* и θ' оптимальны, то

$$M_\theta(\theta_* - \theta)^2 = M_\theta(\theta' - \theta)^2$$

$$M_\theta(\hat{\theta} - \theta)^2 \leq M_\theta(\theta_* - \theta)^2.$$

С другой стороны, так как $\hat{\theta}$ – несмещенная

$$M_\theta(\theta_* - \theta)^2 \leq M_\theta(\hat{\theta} - \theta)^2.$$

Следовательно,

$$M_\theta(\hat{\theta} - \theta)^2 = M_\theta(\theta_* - \theta)^2.$$

Неравенство получилось с помощью неравенства Коши, значит в неравенстве Коши должно быть равенство:

$$M_\theta(\theta_* - \theta)(\theta' - \theta) = \sqrt{M_\theta(\theta_* - \theta)^2}\sqrt{M_\theta(\theta' - \theta)^2}.$$

Если в неравенстве Коши равенство, то функции пропорциональны: $\exists k(\theta) : (\theta_* - \theta) = k(\theta)(\theta' - \theta)$.

Так как

$$M_{\theta}(\theta_* - \theta)^2 = M_{\theta}(\theta' - \theta)^2,$$

то $k^2(\theta) = 1$.

$$\theta_* - \theta = \pm(\theta' - \theta).$$

– 1 оказаться не может.

Упр. Почему?

◁

Пример (оптимальной оценки). $x_1 = \begin{cases} 1, & p \\ 0, & 1-p, \end{cases} \quad p \in [0; 1],$

x_1, \dots, x_n . Оценка

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}.$$

Утверждение. \bar{x} – оптимальная несмещенная оценка

$$M_p(\bar{x} - p)^2 = \frac{1}{n}p(1-p).$$

▷ Пусть \hat{p} – несмещенная оценка для p .

$$M_p(\hat{p} - p)^2 \quad -?$$

$M_p \hat{p} = p$ (т.к. \hat{p} – несмещенная).

$$\sum_{x_1, \dots, x_n \in \{0;1\}} p^{n\bar{x}}(1-p)^{n-n\bar{x}} \cdot \hat{p}(x_1, \dots, x_n) = p$$

$\forall p \in [0; 1]$

$$p^{n\bar{x}}(1-p)^{n-n\bar{x}} = f_{x_1, \dots, x_n}(p).$$

Продифференцировав равенство по $p \quad \forall x_1, \dots, x_n \in \{0; 1\}$

$$\sum_{x_1, \dots, x_n \in \{0;1\}} \frac{f'_{x_1, \dots, x_n}(p)}{f_{x_1, \dots, x_n}(p)} f_{x_1, \dots, x_n}(p) \cdot \hat{p}(x_1, \dots, x_n) = 1.$$

Определим $u(x_1, \dots, x_n) = \frac{f'_{x_1, \dots, x_n}(p)}{f_{x_1, \dots, x_n}(p)}$

$$1 = M_p(u \cdot \hat{p})$$

Докажем, что $M_p u = 0$

$$\sum_{x_1, \dots, x_n} f_{x_1, \dots, x_n}(p) = 1$$

$$\sum_{x_1, \dots, x_n} f'_{x_1, \dots, x_n}(p) = 0 = M_p u$$

$$M_p u(\hat{p} - p) = 1$$

$$\sqrt{M_p(\hat{p} - p)^2} \sqrt{M_p u^2} \geq 1$$

$$\Downarrow$$

$$M_p(\hat{p} - p)^2 \geq \frac{1}{M_p u^2}$$

Упражнение. Посчитать $M_p u^2$. ◁

Достаточные статистики, условные математические ожидания, условные распределения и теорема Колмогорова об улучшении оценок

Пример (улучшения несмещенной оценки).

Рассматриваем

F_θ – непрерывные распределения (функции распределения – непрерывны)
 $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ – несмещенная оценка для θ .

$$\theta_* = \frac{1}{n!} \sum_{\sigma \in S_n} \hat{\theta}(X_{\sigma(1)}, \dots, X_{\sigma(n)})$$

S_n – группа перестановок n элементов.

Докажем, что

1) $M_\theta \theta_* = \theta$ (т.к. $X_{\sigma(1)}, \dots, X_{\sigma(n)}$ – независимые одинаково распределенные и каждый раз распределение одинаково). $\hat{\theta}(x_{\sigma(1)}, \dots, \hat{x}_{\sigma(n)})$

2) $\forall \theta \in \Theta \quad D_\theta \theta_* \leq D_\theta \hat{\theta}$

$\triangleright X_1, \dots, X_n \rightarrow X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ каждую реализацию набора упорядочиваем по возрастанию.

$X_{(1)}, \dots, X_{(n)}$ – вариационный ряд, построенный по выборке.

Если взять 2 перестановки одного набора (X_1, \dots, X_n) , то вариационные ряды, построенные по ним, одинаковы.

$$D_\theta \theta_* = M_\theta \left(\frac{1}{n!} \sum_{\sigma \in S_n} (\hat{\theta}(X_{\sigma(1)}, \dots, X_{\sigma(n)}) - \theta) \right)^2 =$$

$(S_n - \text{группа})$

$$\begin{aligned}
&= M_\theta \left(\frac{1}{n!} \sum_{\varkappa \in S_n} (\hat{\theta}(X_{\varkappa(1)}, \dots, X_{\varkappa(n)}) - \theta) \right)^2 \leq \\
&\leq M_\theta \frac{1}{n!} \sum_{\varkappa \in S_n} (\hat{\theta}(X_{\varkappa(1)}, \dots, X_{\varkappa(n)}) - \theta)^2 = D_\theta \hat{\theta} \quad (*)
\end{aligned}$$

(*) получаем по неравенству

$$\left(\sum_{k=1}^n a_k b_k \right)^2 \leq \left(\sum_{k=1}^n a_k^2 \right) \left(\sum_{k=1}^n b_k^2 \right) \triangleleft$$

Пример. X_1, X_2 – независимые равномерно распределенные на $[0; 1]$

Рис. Сосредоточено плотность вариационного ряда

Интеграл по квадрату равен интегралу по треугольнику, где сосредоточена плотность вариационного ряда плюс интеграл по этому же треугольнику в переставленных координатах, таким образом, симметризованная оценка всегда не хуже.

Мы построили объект (вариационный ряд) такой, что:

$$Mf(X_1, \dots, X_n) = M \left(\frac{1}{n!} \sum_{\varkappa \in S_n} f(X_{\varkappa(1)}, \dots, X_{\varkappa(n)}) \right).$$

Условные математические ожидания

(Ω, \mathcal{F}, P) – вероятностное пространство. Так как не все случайные события можно детектировать, строим

$$\mathcal{F}_1 \subset \mathcal{F}, \mathcal{F}_1 - \text{под } \sigma\text{-алгебра}$$

1. $\mathcal{F}_1 = \{\emptyset, \Omega\}$ лежит в любой σ -алгебре.

2. $\mathcal{F}_1 = \mathcal{F}$

3. $\Omega = \mathbb{R}^n, \mathcal{F}$ – борелевские подмножества \mathbb{R}^n .

\mathcal{F}_1 – борелевские, симметричные относительно всех перестановок.

Упражнение. \mathcal{F}_1 – σ -алгебра.

Пусть ξ – интегр. случайная величина ($M|\xi| < +\infty$).

Определение. Условным математическим ожиданием ξ относительно σ -алгебры \mathcal{F}_1 называется случайная величина η :

1) η – \mathcal{F}_1 -измерима

2) $\forall \Delta \in \mathcal{F}_1 :$

$$M(\xi \cdot \mathbb{I}_\Delta) = M(\eta \cdot \mathbb{I}_\Delta).$$

Обозначается $M(\xi/\mathcal{F}_1)$.

Пример. $\mathcal{F}_1 = (\emptyset, B, \bar{B}, \Omega)$.

$$1 > P(B) > 0$$

$$\xi = \mathbb{I}_A; \quad A \in \mathcal{F}.$$

Все случайные величины, измеримые относительно \mathcal{F}_1 , имеют следующий вид:

$$\eta = C_1 \mathbb{I}_B + C_2 \mathbb{I}_{\bar{B}}$$

$$\Delta = B :$$

$$M \mathbb{I}_A \cdot \mathbb{I}_B = M(C_1 \mathbb{I}_B + C_2 \mathbb{I}_{\bar{B}}) \mathbb{I}_B$$

$$P(A \cap B) = C_1 P(B)$$

$$C_1 = \frac{P(A \cap B)}{P(B)} = P(A/B)$$

$$C_2 = P(A/\bar{B})$$

$$M(\mathbb{I}_A/\mathcal{F}_1) = P(A/B) \mathbb{I}_B + P(A/\bar{B}) \mathbb{I}_{\bar{B}}$$

Лекция 3

Пусть ξ интегрируема с квадратом

$$M|\xi|^2 < +\infty.$$

Свойства условного математического ожидания для случайных величин со вторым моментом

Лемма. Если $M|\xi|^2 < +\infty$, то $M(\xi/\mathcal{F}_1)$ существует.

▷ Рассмотрим $L_2(\Omega, \mathcal{F}, P)$ – гильбертово

$$(\varkappa, \zeta) := M(\varkappa \cdot \zeta)$$

Рассмотрим $\mathcal{K} = \{\varkappa \in L_2, \varkappa \text{ – измеримы относительно } \mathcal{F}_1\}$

\mathcal{K} – подпространство. Действительно,

$$1) \quad a_1, a_2 \in \mathbb{R}; \quad \varkappa_1, \varkappa_2 \in \mathcal{K}$$

$$a_1 \varkappa_1 + a_2 \varkappa_2 \in \mathcal{K}.$$

2) Пусть $\varkappa_n \in \mathcal{K}, n \geq 1$:

$$\varkappa_n \xrightarrow{L_2} \varkappa_0, n \rightarrow \infty$$

Тогда

$$\exists \varkappa_{n_k} \xrightarrow{\text{п.н.}} \varkappa_0, k \rightarrow \infty$$

Отсюда вытекает, что \varkappa_0 – измерима относительно \mathcal{F}_1 с точностью до множества нулевой вероятности.

С этого момента считаем, что все рассматриваемые σ -алгебры содержат все множества вероятности 0.

$$\mathfrak{M} = \left\{ \sum_{k=0}^n c_k \mathbb{I}_{\Delta_k}, n \geq 1, \Delta_k \in \mathcal{F}_1, c_k \in \mathbb{R} \right\}.$$

$$\overline{\mathfrak{M}} = \mathcal{K}.$$

Пусть η – ортогональная проекция ξ на K .

Докажем, что η – искомое условное математическое ожидание ξ относительно \mathcal{F}_1 .

1) выполняется,

2) $\xi - \eta \perp K$

$\forall \zeta \in K$:

$$(\zeta, \xi) = (\zeta, (\xi - \eta) + \eta) = (\zeta, (\xi - \eta)) + (\zeta, \eta) = (\zeta \cdot \eta).$$

$$M(\zeta \cdot \xi) = M(\zeta, \eta)$$

(У.м.о. – проектор на K) $\forall \Delta \in \mathcal{F}_1 \quad \mathbb{I}_{\Delta} \in \mathcal{K}$ и

$$M(\xi \mathbb{I}_{\Delta}) = M(\eta \mathbb{I}_{\Delta}). \quad \triangleleft$$

Лемма. Условное математическое ожидание единственно почти наверное (если существует).

▷ Пусть η_1, η_2 – два у.м.о. ξ относительно \mathcal{F}_1 . Тогда

$\eta_1 - \eta_2$ – измерима относительно \mathcal{F}_1 .

$\forall \Delta \in \mathcal{F}_1$

$$M((\eta_1 - \eta_2) \mathbb{I}_{\Delta}) = M\xi \mathbb{I}_{\Delta} - M\xi \mathbb{I}_{\Delta} = 0.$$

Рассмотрим $\Delta = \{\eta_1 - \eta_2 > 0\} \in \mathcal{F}_1$

$$0 = M(\eta_1 - \eta_2) \mathbb{I}_{\{\eta_1 - \eta_2 > 0\}} \Rightarrow P(\Delta) = 0,$$

также $P\{\eta_1 - \eta_2 < 0\} = 0$. Отсюда вытекает, что

$$\eta_1 = \eta_2 \text{ п.н.} \quad \triangleleft$$

Свойства условных математических ожиданий для случайных величин со вторым моментом (как проектора)

1. Линейность

$$\begin{aligned} M(a_1\xi_1 + a_2\xi_2/\mathcal{F}_1) &= \\ &= a_1M(\xi_1/\mathcal{F}_1) + a_2M(\xi_2/\mathcal{F}_1). \end{aligned}$$

2. Если $\mathcal{F}_1 \subset \mathcal{F}_2$, то

$$M(M(\xi/\mathcal{F}_2)/\mathcal{F}_1) = M(\xi/\mathcal{F}_1)$$

3. Если $\mathcal{F}_1 = \{\emptyset, \Omega\}$, то

$$M(\xi/\mathcal{F}_1) = M\xi$$

4. $M(M(\xi/\mathcal{F}_1)) = M\xi$ (по [2] и [3]).

5. (Проектор непрерывен и его норма равна единице и достигается).

ξ измерима относительно $\mathcal{F}_n \Leftrightarrow M(\xi/\mathcal{F}_n) = \xi$.

ξ не зависит от \mathcal{F}_1 , если

$\forall c \in \mathbb{R} \forall \Delta \in \mathcal{F}_1 : \{\xi \leq c\}$ и Δ независимы. (Обозначаем $(\xi \perp \mathcal{F}_1)$).

6. Если $\xi \perp \mathcal{F}_1$, то

$$M(\xi/\mathcal{F}_1) = M\xi$$

$\triangleright \varkappa$ – произвольная, измеримая относительно \mathcal{F}_1 случайная величина.
Тогда

$$M(\varkappa\xi) = M\varkappa \cdot M\xi.$$

С другой стороны

$$M(\varkappa\xi) = M\varkappa \cdot M(\xi/\mathcal{F}_1)$$

Из единственности у.м.о. следует, что

$$M\xi = M(\xi/\mathcal{F}_1).$$

\triangleleft

7. Пусть ζ – измерима относительно \mathcal{F}_1 , ξ – произвольна. Тогда

$$M(\zeta\xi/\mathcal{F}_1) = \zeta M(\xi/\mathcal{F}_1)$$

$\triangleright \varkappa$ – измеримая относительно \mathcal{F}_1 .

$$M(\varkappa\zeta\xi) = M\varkappa M(\zeta\xi/\mathcal{F}_1)$$

$$M(\varkappa\zeta\xi) = M\varkappa\zeta M(\xi/\mathcal{F}_1)$$

откуда вытекает

$$M(\zeta\xi/\mathcal{F}_1) = \zeta M(\xi/\mathcal{F}_1). \quad \triangleleft$$

Условное математическое ожидание в дискретном случае

Пример. $\mathcal{F}_1 = \sigma(A_1, \dots, A_n)$, где A_1, \dots, A_n – непересекающиеся, $P(A_k) > 0$, $\bigcup_{k=1}^n A_k = \Omega$

η – измерима относительно $\mathcal{F}_1 \Leftrightarrow \eta = \sum_{k=1}^n C_k \mathbb{I}_{A_k}$.

Пусть ξ – интегрируема, как выглядит $M(\xi/\mathcal{F}_1)$?

$$M(\xi/\mathcal{F}_1) = \sum_{k=1}^n X_k \mathbb{I}_{A_k}$$

$$\forall k = \overline{1, n}$$

$$M\xi \mathbb{I}_{A_k} = M \sum_{j=1}^n X_j \mathbb{I}_{A_j} \mathbb{I}_{A_k} = MX_k \mathbb{I}_{A_k} = X_k P(A_k)$$

$$X_k = \frac{1}{P(A_k)} M\xi \mathbb{I}_{A_k} = \frac{1}{P(A_k)} \int_{A_k} \xi(\omega) P(d\omega)$$

то есть у.м.о. сводится к интегрированию. По аналогии с интегрированием получаем свойства:

8. φ – выпуклая вниз функция. Тогда

$$\varphi(M(\xi/\mathcal{F}_1)) \leq M(\varphi(\xi)/\mathcal{F}_1) \text{ п.н.}$$

В частном случае

$$|M(\xi/\mathcal{F}_1)| \leq M(|\xi|/\mathcal{F}_1).$$

9. а) $\xi_n \geq 0, \xi_n \nearrow \xi, n \rightarrow \infty$, тогда

$$M(\xi_n/\mathcal{F}_1) \nearrow M(\xi/\mathcal{F}_1)$$

б) $\xi_n \xrightarrow{P} \xi, n \rightarrow \infty, |\xi_n| \leq \eta, M(\eta) < +\infty$, тогда

$$M(\xi_n/\mathcal{F}_1) \rightarrow M(\xi/\mathcal{F}_1) \text{ п.н.}$$

10. $\xi \geq 0$ п.н. $\Rightarrow M(\xi/\mathcal{F}_1) \geq 0$ п.н.

Предложение. 1. УМО существует для случайной величины, имеющей первый момент.

2. Все перечисленные свойства сохраняются для случайной величины с первым моментом.

Вычисление УМО

Определение. Совокупность случайных событий вида $\xi^{-1}(A)$, где $A \in \mathcal{B}_{\mathbb{R}}$ называется σ -алгеброй, порожденной случайной величиной ξ и обозначается $\sigma(\xi)$.

Пример.

1. $\Omega = [0; 1]^2$; $\mathcal{B}_{\mathbb{R}}$

$$\xi(\omega_1, \omega_2) = \omega_1$$

Рис.

Множества из $\sigma(\xi)$ имеют вид

2. $\Omega = \mathbb{R}^2$

$$\xi(\omega_1, \omega_2) = \omega_1^2 + \omega_2^2$$

Рис.

Аналогично для $\sigma(\vec{\xi})$.

Пусть η, ξ – случайные величины. Как выглядит $M(\eta/\sigma(\xi))$?

Теорема. Пусть ζ – случайная величина, измеримая относительно $\sigma(\xi)$. Тогда существует борелевская функция $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ такая, что

$$\zeta = \varphi(\xi). \quad \triangleright \triangleleft$$

Следовательно,

$$M(\eta/\sigma(\xi)) = \varphi(\xi).$$

Как найти φ – ?

$$\varphi(x) := M(\eta/\xi = x)$$

\forall функции h

$$M\eta h(\xi) = M\varphi(\xi)h(\xi)$$

1. Случайные величины ξ и η принимают не более чем счетное число значений.

$$\{a_i\}, \{b_j\}$$

$$P\{\xi = a_i, \eta = b_j\} = p_{ij} > 0$$

$$\sum_{i,j} p_{ij} = 1$$

$$M\eta h(\xi) = \sum_{i,j} p_{ij} b_j h(a_i) =$$

$$P\{\xi = a_i\} = \sum_j p_{ij}$$

$$M\varphi(\xi)h(\xi) = \sum_i \varphi(a_i)h(a_i) \left(\sum_j p_{ij} \right)$$

$$\sum_{i,j} b_j h(a_i) p_{ij} = \sum_i \varphi(a_i) h(a_i) \sum_j p_{ij}$$

h – любое.

Возьмем $h_i(x) = \begin{cases} 0, & x \neq a_i \quad \forall i \\ 1, & x = a_i \end{cases}$

$\forall i :$

$$\sum_j b_j p_{ij} = \varphi(a_i) \sum_j p_{ij}$$

$$\varphi(a_i) = \frac{\sum_j b_j p_{ij}}{\sum_j p_{ij}}$$

$$M(f(\eta)/\xi = a_i) = \frac{\sum_j b_j p_{ij}}{\sum_j p_{ij}}$$

Рис.

2. ξ и η имеют совместную плотность распределения $p(x, y)$, $x, y \in \mathbb{R}$

$$M\eta h(\xi) = \iint_{\mathbb{R}^2} y h(x) p(x, y) dx dy$$

$M\varphi(\xi)h(\xi) - ?$

плотность $\xi :$

$$q(x) = \int_{\mathbb{R}} p(x, y) dy$$

$$M\varphi(\xi)h(\xi) = \int_{\mathbb{R}} [\varphi(x)] h(x) \left[\int_{\mathbb{R}} p(x, y) dy \right] dx$$

$$M\eta h(\xi) = \int_{\mathbb{R}} h(x) \left[\int_{\mathbb{R}} y p(x, y) dy \right] dx$$

$$\varphi(x) = \frac{\int_{\mathbb{R}} y p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy}$$

$$M(f(\eta)/\xi = x) = \frac{\int_{\mathbb{R}} f(y) p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy}$$

Лекция 4

\vec{x} – случайный вектор в \mathbb{R}^d (выборка)

$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

$$\xi = f(\vec{x})$$

$$M(\varphi(\vec{x})/\xi = t) - ?$$

Условное математическое ожидание функции от случайного вектора при известной функции от вектора

Пусть p – плотность распределения \vec{x}

$$\begin{aligned} Mh(\xi)\varphi(\vec{x}) &= Mh(f(\vec{x}))\varphi(\vec{x}) = \\ &= \int_{\mathbb{R}^d} h(f(\vec{u}))\varphi(\vec{u})p(\vec{u})d\vec{u}. \end{aligned}$$

Пусть $f \in C^1(\mathbb{R}^d)$ и возможно отобразить $\vec{u} \xrightarrow{F} (f(\vec{u}), v_2, \dots, v_d)$ так, что F – диффеоморфизм (биекция, невырожденная в каждой точке). В этих предположениях плотность распределения ξ

$$\begin{aligned} Mh(\xi) &= Mh(f(\vec{x})) = \int_{\mathbb{R}^d} h(f(\vec{u}))p(\vec{u})d\vec{u} = \\ &= \int_{\mathbb{R}^d} h(v_1)p(F^{-1}(\vec{v}))J_F(\vec{v})d\vec{v} = \\ &= \int_{\mathbb{R}} h(v_1) \underbrace{\left\{ \int_{\mathbb{R}^{d-1}} p(F^{-1}(\vec{v}))J_F(\vec{v})dv_2 \dots dv_d \right\}}_{\text{плотность } \xi, \text{ обозначим ее } q_\xi} dv_1 \end{aligned}$$

Упражнение. При другой замене переменных (v'_1, \dots, v'_d) плотность ξ не меняется (только функция $v_1 = v'_1 = f(u)$)

$$\begin{aligned} Mh(\xi)\varphi(\vec{x}) &= \int_{\mathbb{R}^d} h(f(\vec{u}))\varphi(\vec{u})p(\vec{u})d\vec{u} = \\ &= \int_{\mathbb{R}} h(v_1) \left\{ \int_{\mathbb{R}^{d-1}} \varphi(F^{-1}(\vec{v}))p(F^{-1}(\vec{v}))J_F(\vec{v})dv_2 \dots dv_d \right\} \times \\ &\quad \times q_\xi^{-1}(v_1)q_\xi(v_1)dv_1 \\ g(v_1) &= \frac{\int_{\mathbb{R}^{d-1}} \varphi(F^{-1}(\vec{v}))p(F^{-1}(\vec{v}))J_F(\vec{v})dv_2 \dots dv_d}{\int_{\mathbb{R}^{d-1}} p(F^{-1}(\vec{v}))J_F(\vec{v})dv_2 \dots dv_d} \end{aligned}$$

Пример.

1. $d = 2$ $\vec{x} = (x_1, x_2)$, x_1, x_2 – н.о.р.

$$x_1 \stackrel{d}{=} x_2 \stackrel{d}{=} N(0; 1)$$

$$f(x_1, x_2) = \sqrt{x_1^2 + x_2^2}, \quad \xi = f(x_1, x_2)$$

$$\vec{u} = (u_1, u_2)$$

$$p(\vec{u}) = \frac{1}{2\pi} e^{-\frac{u_1^2 + u_2^2}{2}}$$

Замена: $(\rho = \sqrt{u_1^2 + u_2^2}, \varphi = \operatorname{arctg} \frac{u_1}{u_2})$

$$J(\rho, \varphi) = \rho$$

$$\mathbb{R}^2 \rightarrow \mathbb{R}_+ \times [0; 2\pi]$$

$$\begin{aligned} g(\rho) &= \left(\int_0^{2\pi} \Phi(\rho \cos \varphi, \rho \sin \varphi) \frac{1}{2\pi} e^{-\frac{\rho^2}{2} \rho d\varphi} \right) \div \\ &\div \left(\frac{1}{2\pi} e^{-\frac{\rho^2}{2} \rho d\varphi} \right) = \\ &= \frac{1}{2\pi} \int_0^{2\pi} \Phi(\rho \cos \varphi, \rho \sin \varphi) \varphi = \\ &= M(\Phi(\vec{x}) \sqrt{x_1^2 + x_2^2} = \rho) \end{aligned}$$

Интегрирование производится по окружности, т.е. при вычислении условного математического ожидания интегрирование производится по поверхности уровня $f(\vec{x})$.

Рис.

Условные распределения и достаточные статистики

Пример. $\zeta \sim \operatorname{Pois}(\lambda)$ ($\lambda > 0$)

$\{\xi_k, k \geq 1\}$ – н.о.р.

$\eta = \sum_{k=1}^{\zeta} \xi_k$; $\eta = 0$, если $\zeta = 0$.

ζ и $\{\xi_k\}$ – независимы. Найти характеристическую функцию η .

$Me^{it\eta}$ – ?

$$Me^{it\eta} = M(M(e^{it\eta}/\zeta))$$

$$M(e^{it\eta}/\zeta) = ?$$

$$\zeta = 0, 1, \dots, k, \dots$$

$$P(\zeta = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$M(e^{it\eta} / \zeta = k) = (\varphi_\xi(t))^k$$

$$M(e^{it\eta} / \zeta) = (\varphi_\xi(t))^\zeta$$

$$\begin{aligned} Me^{it\eta} &= M(M(e^{it\eta} / \zeta)) = M((\varphi_\xi(t))^\zeta) = \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda \varphi_\xi(t))^k}{k!} = e^{\lambda(\varphi_\xi(t)-1)} \end{aligned}$$

Пусть $\xi_1 = \begin{cases} 1, & p \\ 0, & 1-p = q \end{cases}$

$$\varphi_\xi(t) = e^{it}p + q = p(e^{it} - 1) + 1$$

$$Me^{it\eta} = e^{\lambda p(e^{it}-1)},$$

и η – пуассоновская с параметром λp .

Определение. Отображение $\mu : \Omega \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0; 1]$ – случайная вероятностная мера, если

- 1) $\forall \omega \in \Omega \quad \mu(\omega; \cdot)$ – вероятностная мера;
- 2) $\forall \Delta \in \mathcal{B}(\mathbb{R}^d) \quad \mu(\cdot; \Delta)$ – случайная величина.

Определение. Условное распределение случайной величины ξ при условии η – это случайная мера μ такая, что

- 1) μ измерима относительно η ,
- 2) для каждой ограниченной измеримой функции φ :

$$M(\varphi(\xi) / \eta) = \int_{\mathbb{R}} \varphi(u) \mu(du)$$

(измерима, так как $\mu(\cdot, du)$ – измерима).

Пример 1. ξ, η имеют совместное дискретное распределение

$$P\{\xi = a_i, \eta = b_j\} = p_{ij}$$

$$M(\varphi(\xi) / \zeta = b_j) = \frac{\sum_i \varphi(a_i) p_{ij}}{\sum_i p_{ij}}$$

Найдем условное распределение.

Условное распределение ξ при известном η – вероятностная мера, сосредоточенная в точках a_i с весами $\frac{p_{ij}}{\sum_i p_{ij}}$.

Рис.

2. ξ и η имеют совместную плотность распределения $p(u_1, u_2)$

$$\begin{aligned} M(\varphi(\xi)/\eta = y) &= \\ &= \frac{\int_{\mathbb{R}} \varphi(x) p(x, y) dx}{\int_{\mathbb{R}} p(x, y) dx} \\ \mu(dx) &= \frac{p(x, \eta)}{\int_{\mathbb{R}} p(x, \eta) dx} dx. \end{aligned}$$

3. \vec{x} , $\xi = f(\vec{x})$.

Условное распределение \vec{x} при известном ξ

$$M(\varphi(\vec{x})/\xi = v_1) = \frac{\int_{\mathbb{R}^{d-1}} \varphi(F^{-1}(\vec{v})) p(F^{-1}(\vec{v})) J_F(\vec{v}) dv_2 \dots dv_d}{\int_{\mathbb{R}^{d-1}} p(F^{-1}(\vec{v})) J_F(\vec{v}) dv_2 \dots dv_d}$$

Из примера

Рис.

$$\frac{1}{2\pi} \int_0^{2\pi} \Phi(\rho \cos \varphi, \rho \sin \varphi) d\varphi = \int_{\Gamma_\rho} \Phi(\vec{v}) \frac{\sigma(d\vec{v})}{2\pi\rho}$$

σ – пов. мера на окружности.

Определение. Функция от выборки $T(\vec{x})$ называется достаточной статистикой, если условное распределение \vec{x} при известном $T(\vec{x})$ не зависит от неизвестного параметра θ .

Пример.

1. Выборка из непрерывного распределения.

$T(\vec{x}) = (x_{(1)}, \dots, x_{(n)})$ – вариационный ряд.

$$M(\varphi(\vec{x})/T(\vec{x})) = h(T(\vec{x}))$$

$$M\varphi(\vec{x})g(T(\vec{x})) = Mh(T(\vec{x}))g(T(\vec{x}))$$

x_1 имеет плотность распределения $p_\theta(t)$.

\vec{x} имеет распределение $p_\theta(u_1) \dots p_\theta(u_n)$

$T(\vec{x})$ имеет плотность распределения

$$n! p_\theta(u_1), \dots, p_\theta(u_n) \mathbb{I}_{\{u_1 \leq u_2 \leq \dots \leq u_n\}} = q_\theta(\vec{u})$$

$$M\varphi(\vec{x})g(T(\vec{x})) = \int_{\varphi_1 \leq \dots \leq \varphi_n} q_\theta(\vec{u}) g(\vec{u}) \frac{1}{n!} \sum_{\sigma \in S_n} \varphi(\varphi_{\sigma(1)}, \dots, \varphi_{\sigma(n)}) d\vec{u}.$$

Условное распределение \vec{x} при известном вариационном ряде

$$\frac{1}{n!} \sum_{\sigma \in S_n} \delta(u_{\sigma(1)}, \dots, u_{\sigma(n)}),$$

где u_1, \dots, u_n – члены вариационного ряда.

Нет никакой зависимости от параметра θ .

2. \vec{x} – выборка из $Pois(\lambda)$

$S = \sum_{k=1}^n X_k$ – достаточная статистика для λ .

▷

$$M(\varphi(\vec{x})/S = s)$$

$$x_1 = k_1, \dots, x_n = k_n$$

$$\sum_{i=1}^n k_i = s$$

$$e^{-\lambda_n} \frac{\lambda^{k_1} \dots \lambda^{k_n}}{k_1! \dots k_n!} = e^{-\lambda_n} \frac{\lambda^s}{k_1! \dots k_n!}$$

$$M(\varphi(\vec{x})/S = s) = \frac{\sum_{k_1+\dots+k_n=s} \frac{\varphi(k_1, \dots, k_n)}{k_1! \dots k_n!}}{\sum_{k_1+\dots+k_n=s} \frac{1}{k_1! \dots k_n!}}$$

Значения φ складываются с весами

$$\frac{1}{k_1! \dots k_n!} \left(\sum_{k_1+\dots+k_n=S} \frac{1}{k_1! \dots k_n!} \right)^{-1}. \quad \triangleleft$$

Лекция 5

$$F_\theta, \theta \in \Theta$$

$$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$$

T – достаточная статистика.

$M(\hat{\theta}/T) = \theta^*$ не зависит от θ .

Теорема (характеризация для достаточной статистики.) Пусть $\forall \theta \in \Theta$ распределение F_θ имеет плотность p_θ . Статистика T является достаточной тогда и только тогда, когда p_θ можно представить в виде

$$p_\theta(\vec{x}) = m(\vec{x})h(T, \theta).$$

Замечание. Утверждение теоремы остается справедливым в том случае, когда все рассматриваемые распределения сосредоточены на множестве целых чисел.

▷ \Rightarrow Пусть T – гладкая функция от \vec{x} . Условное распределение \vec{x} при известном T сосредоточены на поверхностях $T = t, t \in \mathbb{R}$.

$$\begin{aligned}
M_\theta(f(\vec{x})) &= \int_{\mathbb{R}^n} f(\vec{u}) p_\theta(\vec{u}) d\vec{u} = \\
&= M_\theta(M_\theta(f(\vec{x})/T)) = \\
&= \int_{\mathbb{R}} q(\theta, t) \left[\int_{\{T=t\}} f(\vec{u}) g(\vec{u}, t) \sigma_t(d\vec{u}) \right] dt = \\
&\quad (t, u_2, \dots, u_n) \leftrightarrow (x_1, \dots, x_n) \\
&= \int_{\mathbb{R}^n} f(\vec{x}) q(\theta, T(\vec{x})) m(\vec{x}) d\vec{x}
\end{aligned}$$

$$\forall f \Rightarrow p_\theta(\vec{x}) = m(\vec{x}) q(\theta, T(\vec{x})).$$

Если статистика T – достаточна.

\Leftarrow Пусть $p_\theta(\vec{x}) = m(\vec{x}) q(\theta, T(\vec{x}))$. Тогда при замене переменных условное распределение не будет зависеть от θ . \triangleleft

Замечание. Теорема справедлива во всех случаях на практике.

Примеры (достаточных статистик)

1. $Exp(\lambda); p_\lambda(x) = \lambda e^{-\lambda x}, x \geq 0$
 $\lambda > 0$

$$p_\lambda(\vec{x}) = \lambda^n e^{-\lambda(x_1 + \dots + x_n)} \prod_{k=1}^n \mathbb{I}_{[0; +\infty)}(x_k)$$

$$p_\lambda(\vec{x}) = m(\vec{x}) h(T(\vec{x}), \lambda)$$

$$T(\vec{x}) = x_1 + \dots + x_n$$

$$h(T, \lambda) = \lambda^n e^{-\lambda T}$$

$$m(\vec{x}) = \prod_{k=1}^n \mathbb{I}_{[0; +\infty)}(x_k)$$

T – достаточная статистика.

2. Равномерное распределение на $[0; \theta], \theta \in (0; +\infty)$

$$p_\theta(x) = \frac{1}{\theta} \mathbb{I}_{[0; \theta]}(x)$$

$$p_\theta(\vec{x}) = \frac{1}{\theta^n} \prod_{k=1}^n \mathbb{I}_{[0; +\infty)}(x_k) = \frac{1}{\theta^n} \prod_{k=1}^n \mathbb{I}_{[0; +\infty)}(x_k) \mathbb{I}_{[0; \theta]}(\max_{k=1, \dots, n} x_k) =$$

$$= \prod_{k=1}^n \mathbb{I}_{[0;+\infty)}(x_k) \frac{1}{\theta^n} \mathbb{I}_{[0;\theta)}(\max_{k=\overline{1,n}} x_k).$$

$T = \max_{k=\overline{1,n}} x_k$ – не гладкая достаточная статистика.

Упражнение. Найти достаточную статистику для θ в распределении Коши:

$$p_\theta(x) = \frac{1}{n} x \frac{1}{1 + (x - \theta)^2}$$

$\hat{\theta}$ – несмещенная, T – достаточная статистика.

Лемма. $\forall \theta \in \Theta$

$$D_\theta \theta^* \leq D_\theta \hat{\theta},$$

где $\theta^* = M(\hat{\theta}/T)$

▷

$$D_\theta \theta^* = M_\theta(M(\hat{\theta}/T) - \theta)^2 =$$

(неравенство Йенсена, или Коши–Буняковского)

$$= M_\theta(M(\hat{\theta} - \theta/T))^2 \leq M_\theta M((\hat{\theta} - \theta)^2/T) =$$

$$= D_\theta \hat{\theta}. \quad \triangleleft$$

Оптимальная оценка – функция какой-то достаточной статистики.

Оценки максимального правдоподобия и неравенство Рао–Крамера

F_θ имеет плотность p_θ , которую можно дважды дифференцировать по θ , меняя дифференц. и математическое ожидание местами.

Замечание. Все рассуждения остаются справедливыми, когда речь идет о распределении, сосредоточенном на множестве целых чисел.

p_θ – одномерная плотность,

$L(\vec{x}, \theta)$ – многомерная плотность выборки

$L(\vec{x}, \theta) = \prod_{k=1}^n p_\theta(x_k)$ – функция правдоподобия.

$U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta)$ – вклад выборки.

$M_\theta U^2(\vec{x}, \theta) = I_n(\theta)$ – количество информации, содержащейся в выборке.

Свойства вклада выборки:

1. $M_\theta(U(\vec{x}, \theta)) = 0$

▷

$$M_\theta(U(\vec{x}, \theta)) = \int_{\mathbb{R}^d} U(\vec{x}, \theta) L(\vec{x}, \theta) d\vec{x}$$

$$\int_{\mathbb{R}^n} L(\vec{x}, \theta) d\vec{x} = 1$$

(дифф. по θ)

$$\begin{aligned} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\vec{x}, \theta) d\vec{x} &= 0 \\ \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \frac{L(\vec{x}, \theta)}{L(\vec{x}, \theta)} L(\vec{x}, \theta) d\vec{x} &= 0. \end{aligned}$$

◁

2.

$$\begin{aligned} M_\theta U(\vec{x}, \theta)^2 &= -M_\theta \frac{\partial}{\partial \theta} U(x, \theta) = \\ &= -M_\theta \frac{\partial^2}{\partial \theta^2} \ln L(x, \theta). \end{aligned}$$

▷

$$\begin{aligned} \int_{\mathbb{R}^n} U(\vec{x}, \theta) L(\vec{x}, \theta) d\vec{x} &= 0 \\ \int_{\mathbb{R}^n} U(\vec{x}, \theta) \frac{\partial}{\partial \theta} L(\vec{x}, \theta) d\vec{x} + \\ + \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} (U(\vec{x}, \theta)) L(\vec{x}, \theta) d\vec{x} &= 0 \\ \int_{\mathbb{R}^n} U(\vec{x}, \theta)^2 L(\vec{x}, \theta) d\vec{x} &= \\ = - \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} U(\vec{x}, \theta) L(\vec{x}, \theta) d\vec{x} \\ M_\theta U(\vec{x}, \theta)^2 &= -M_\theta \frac{\partial}{\partial \theta} U(x, \theta). \quad \triangleleft \end{aligned}$$

3. $I_n(\theta) = M_\theta U(\vec{x}, \theta)^2$

$$U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} L(x, \theta) = \sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p_\theta(x_k)$$

$$I_n(\theta) = M_\theta \left(\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p_\theta(x_k) \right)^2 =$$

(x_k – незав. о.р. и $M_\theta \frac{\partial}{\partial \theta} \ln p_\theta(x_k) = 0$)

$$= \sum_{k=1}^n M_\theta \left(\frac{\partial}{\partial \theta} \ln p_\theta(x_k) \right)^2 = M_\theta \left(\frac{\partial}{\partial \theta} \ln p_\theta(x_k) \right)^2$$

$I_n(\theta) = nI(\theta)$, $I(\theta)$ – количество информации, отвечающее выборке объема 1.

Пример (вычисление количества информации).

1. *Gauss*($\theta, 1$)

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$$

$\theta \in \mathbb{R}$

$$\ln p_\theta(x) = \ln \frac{1}{\sqrt{2\pi}} - \frac{(x-\theta)^2}{2}$$

$$\frac{\partial^2}{\partial \theta^2} \ln p_\theta(x) = -1, \quad I_n(\theta) = -n$$

Gauss(θ, σ^2)

$$I(\theta) = -\frac{1}{\sigma^2}.$$

2. *Cauchy*(θ)

$$p_\theta(x) = \frac{1}{n} \frac{1}{1 + (x - \theta)^2}$$

$$\ln p_\theta(x) = \ln \frac{1}{n} - \ln(1 + (x - \theta)^2)$$

$$\frac{\partial}{\partial \theta} \ln p_\theta(x) = + \frac{2(x - \theta)}{1 + (x - \theta)^2}$$

$$\frac{\partial^2}{\partial \theta^2} \ln p_\theta(x) = \frac{-2(1 + (x - \theta)^2) + 4(\theta - x)^2}{1 + (x - \theta)^2}$$

Интеграл с помощью вычетов.

Неравенство Рао–Крамера

Теорема. Пусть $\hat{\theta}$ – несмещенная оценка θ . Тогда

$$D_\theta \hat{\theta} \geq \frac{1}{I_n(\theta)}$$

$\triangleright \hat{\theta}$ – несмещенная $\Rightarrow M_\theta \hat{\theta} = \theta$

$$M_\theta \hat{\theta} U(x, \theta) = 1$$

$$M_\theta (\hat{\theta} - \theta) U(x, \theta) = 1$$

$$1 \leq \sqrt{M_\theta (\hat{\theta} - \theta)^2} \sqrt{M_\theta U(x, \theta)^2}. \quad \triangleleft$$

Упражнение. Если $M_{\theta}\hat{\theta} = f(\theta)$, то неравенство Рао–Крамера имеет вид:

$$D_{\theta}\hat{\theta} \geq \frac{(f'(\theta))^2}{I_n(\theta)}$$

Определение. Оценка максимального правдоподобия θ^* :

$$\theta^* = \arg \max_{\theta \in \Theta} L(\vec{x}, \theta)$$

(т.е. такое θ^* , что $L(\vec{x}, \theta^*) = \max_{\theta \in \Theta} L(\vec{x}, \theta)$)

Лекция 6

Свойства оценок максимального правдоподобия

Замечание. Пусть T – достаточная статистика, $\hat{\theta}$ – несмещенная оценка.

Тогда $\tilde{\theta} = M(\hat{\theta}/T)$ – улучшенная оценка.

1. Если θ^* выбирается единственным образом и T – достаточная статистика, то θ^* – функция от T .

▷ По теор. характеристики для достаточных статистик

$$L(\theta, \vec{x}) = m(\vec{x})h(\theta, T)$$

$$\theta^* = f(T). \quad \triangleleft$$

Свойства УМО:

1. Если $M(\xi/T) = \zeta$, ζ – измерима относительно $\sigma(T)$. Значит $\zeta = \varphi(T)$.

2. Если ξ измерима относительно T , то

$$M(\xi/T) = \xi.$$

Следовательно,

$$\tilde{\theta} = M(\theta^*/T) = \theta^*$$

(θ^* измерима относительно T .)

$$U(\theta, \vec{x}) = \frac{\partial}{\partial \theta} \ln L(\theta, \vec{x})$$

$$I(\theta) = M_{\theta}U^2(\theta, \vec{x}) = nI_1(\theta)$$

Если $\hat{\theta}$ – несмещенная оценка для θ , то

$$D_{\theta}\hat{\theta} \geq \frac{1}{I_n(\theta)}$$

Определение. Оценка $\hat{\theta}$, для которой $D_{\theta}\hat{\theta} = \frac{1}{I_n(\theta)}$, $\forall \theta \in \Theta$ называется эффективной.

Эффективная оценка всегда оптимальная.

Когда существует эффективная оценка?

$$\begin{aligned} 1 &= M_{\theta}[(\hat{\theta} - \theta)U(\theta, \vec{x})] \leq \\ &\leq \sqrt{D_{\theta}\hat{\theta}I_n(\theta)} \quad (\text{неравенство Коши}) \end{aligned}$$

В неравенстве Коши равенство достигается, когда сомножители слева пропорциональны:

$$(\hat{\theta} - \theta)K(\theta) = U(\theta, \vec{x})$$

$\hat{\theta} = \hat{\theta}(\vec{x})$ – некоторая функция от выборки.

$$U_{\theta}(\theta, \vec{x}) = \frac{\partial}{\partial \theta} \ln L(\theta, \vec{x})$$

$$\ln L(\theta, \vec{x}) = A(\theta) + B(\theta)f(\vec{x}) + g(\vec{x})$$

$g(\vec{x})$ – константа, возникающая при интегрировании

$$L(\theta, \vec{x}) = \exp\{B(\theta)f(\vec{x}) + A(\theta) + g(\vec{x})\} \quad (*)$$

– вид функции правдоподобия для существования эффективной оценки.

Определение. Семейства, для которых функция правдоподобия имеет вид $(*)$ называются экспоненциальными.

Примеры.

1. $N(\theta, 1)$

$$\begin{aligned} p_{\theta}(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}, \quad \theta \in \mathbb{R} \\ L(\theta, \vec{x}) &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{k=1}^n (x_k - \theta)^2} = \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{k=1}^n x_k^2 + \theta \sum_{k=1}^n x_k - \frac{1}{2} n \theta^2} \end{aligned}$$

– вид $(*)$

2. $N(\theta, \sigma^2); (\theta, \sigma^2) \in \mathbb{R} \times (0; +\infty)$

$$L(\theta, \sigma^2, \vec{x}) = \frac{1}{(2\pi)^{n/2} \sigma^{2n/2}} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \theta)^2} =$$

$$\begin{aligned}
&= \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\ln \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2 + \frac{\theta}{\sigma^2} \sum_{k=1}^n x_k - \frac{n\theta^2}{2\sigma^2} \right\} = \\
&= \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2 + \frac{\theta}{\sigma^2} \sum_{k=1}^n x_k - \frac{n}{2} (\ln \sigma^2 + \frac{\theta^2}{\sigma^2}) \right\}.
\end{aligned}$$

Упражнение. Довести до конца.

3. $Exp(\lambda)$; $\lambda \in (0; +\infty)$

$$p_\lambda(x) = \lambda e^{-\lambda x} \mathbb{I}_{\{x>0\}}$$

$$\begin{aligned}
L(\lambda, \vec{x}) &= \lambda^n e^{-\lambda \sum_{k=1}^n x_k} \mathbb{I}_{\{\min_k x_k > 0\}} = \\
&= \exp \{ n \ln \lambda - \lambda \sum_{k=1}^n x_k + \ln \mathbb{I}_{\{0; +\infty\}}(\min \lambda) \}
\end{aligned}$$

4. $Pois(\lambda)$, $\lambda > 0$

$$\begin{aligned}
L(\lambda, \vec{x}) &= e^{-n\lambda} \prod_{k=1}^n \frac{\lambda^{x_k}}{(x_k)!} = \left| x_k \geq 0, x_k \in \mathbb{Z}, k = 1, \dots, n \right| = \\
&= \exp \{ -n\lambda + \ln(\lambda \sum_{k=1}^n x_k) - \sum_{k=1}^n \ln(x_k!) \}
\end{aligned}$$

Как искать оценку максимального правдоподобия?

$$\theta^* = \arg \max_{\theta \in \Theta} L(\theta, \vec{x})$$

Рассматривать патологические случаи не будем, т.е. предполагаем, что \max всегда существует и единственный

$$\theta^* = \arg \max_{\theta \in \Theta} L(\theta, \vec{x})$$

Ищем максимум.

Для θ^* **уравнение правдоподобия:**

$$\frac{\partial}{\partial \theta} \ln L(\theta, \vec{x}) = 0 \Leftrightarrow$$

$$U(\theta, \vec{x}) = 0$$

Для экспоненциального семейства

$$L(\theta, \vec{x}) = \exp \{ A(\theta) + B(\theta) f(\vec{x}) + g(\vec{x}) \}$$

$$\ln L(\theta, \vec{x}) = A(\theta) + B(\theta)f(\vec{x}) + g(\vec{x})$$

Уравнение правдоподобия:

$$A'(\theta) + f(\vec{x})B'(\theta) = 0$$

Пусть θ^* – решение уравнения правдоподобия

$$A'(\theta^*) + B'(\theta^*)f(\vec{x}) = 0$$

$$f(\vec{x}) = -\frac{A'(\theta^*)}{B'(\theta^*)}$$

Для равенства в неравенстве Рао–Крамера:

$$U(\theta, \vec{x}) = K(\theta)(\theta^* - \theta)$$

$$A'(\theta) + B'(\theta)f(\vec{x}) = A'(\theta) - B'(\theta)\frac{A'(\theta^*)}{B'(\theta^*)}$$

$$K(\theta)\theta^* - \theta K(\theta) = -B'(\theta)\frac{A'(\theta^*)}{B'(\theta^*)} + A'(\theta)$$

Пусть

$$B'(\theta) = a\theta + b \rightarrow B(\theta) = a\theta + b$$

$$A'(\theta) = c\theta + d$$

$$-B'(\theta)\frac{A'(\theta^*)}{B'(\theta^*)} + A'(\theta) = -a\frac{c\theta^* + d}{a} + c\theta + d =$$

$$= -c\theta^* - d + c\theta + d = c(\theta - \theta^*) = -c(\theta^* - \theta), \quad K(\theta) = c.$$

Тогда равенство выполняется.

Когда $B(\theta)$ – линейная, а $A(\theta)$ – квадратичная, то о.м.п. – эффективная.

Уравнение правдоподобия:

$$U(\theta, \vec{x}) = 0$$

$$\sum_{k=1}^n \underbrace{\frac{\partial}{\partial \theta} \ln p(\theta, x_k)}_{\xi_k} = 0,$$

тогда $\{\xi_k, k \geq 1\}$ – н.о.р.

$$M_\theta \xi_k = 0$$

$$D_\theta \xi_k = I_1(\theta)$$

Используем Центральную предельную теорему

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(\theta, x_k) \xrightarrow[n \rightarrow \infty]{} N(0, I_1(\theta))$$

$$\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(\theta^*, x_k) = 0.$$

По формуле Тейлора

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln p(\theta^*, x_k) &= \frac{\partial}{\partial \theta} \ln p(\theta, x_k) + \frac{\partial^2}{\partial \theta^2} \ln p(\theta, x_k)(\theta^* - \theta) + \\ &+ \frac{1}{2} \frac{\partial^3}{\partial \theta^3} \ln p(\theta, x_k)(\theta^* - \theta)^2 + r(\theta^*) \end{aligned}$$

Суммируем по $k = \overline{1, n}$, делим на n и переходим к пределу при $n \rightarrow \infty$.

$$0 = (\theta^* - \theta)I_1(\theta) + \frac{(\theta^* - \theta)^2}{2} M_\theta \frac{\partial^3}{\partial \theta^3} \ln p(\theta, x_1)$$

Линейные оценки и гауссовские системы

Примеры. 1. $N(\theta, 1)$

$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ – эффективная оценка для θ

\bar{x} – оптимальная $\Rightarrow \bar{x}$ – оптимальная среди всех линейных оценок

$\hat{\theta} = \sum_{k=1}^n a_k x_k$ – линейная оценка.

Задача поиска наилучшей линейной оценки

$$\begin{aligned} M_\theta(\hat{\theta} - \theta)^2 &= M_\theta \left(\sum_{k=1}^n a_k x_k - \theta \right)^2 = \\ &= M_\theta \sum_{k_1, k_2=1}^n a_{k_1} a_{k_2} x_{k_1} x_{k_2} - 2\theta M_\theta \sum_{k=1}^n a_k x_k + \theta^2 = \\ &= \sum_{k_1, k_2=1}^n a_{k_1} a_{k_2} M_\theta x_{k_1} x_{k_2} - 2\theta \sum_{k=1}^n a_k x_k + \theta^2 \\ M_\theta x_1 &= m_1(\theta) \\ M_\theta x_{k_1} x_{k_2} &= \begin{cases} m_1(\theta)^2, & k_1 \neq k_2 \\ m_2(\theta), & k_1 = k_2 \end{cases} \end{aligned}$$

Минимум ищется дифференцированием.

Лекция 7

Оптимальная линейная оценка будет определяться только первыми двумя моментами. Следовательно, поведение оптимальной линейной оценки будет таким же, как и для $N(m_1(\theta), m_2(\theta))$.

Гауссовские случайные вектора

\mathbb{R}^d ; $\vec{\xi} = (\xi_1, \dots, \xi_d)$

Определение. ξ – гауссовская случайная величина, если $\forall \vec{\varphi} \in \mathbb{R}^d$:

$$(\vec{\xi}, \vec{\varphi}) = \sum_{i=1}^d \xi_i \varphi_i$$

– гауссовская случайная величина.

Замечание. При выполнении условия определения, случайные величины ξ_1, \dots, ξ_d называются совместно гауссовскими.

Пример. 1. $\vec{\xi} = (\eta, \dots, \eta); \eta \sim N(a, \sigma^2)$

$\vec{\xi}$ – гауссовский случайный вектор.

$$(\vec{\xi}, \vec{\varphi}) = \left(\sum_{i=1}^d \varphi_i \right) \eta \quad \forall \vec{\varphi} \in \mathbb{R}^d.$$

2. Стандартный гауссовский вектор.

$\vec{\xi} = (\xi_1, \dots, \xi_d)$, ξ_1, \dots, ξ_d – н.о.р. $\xi_1 \sim N(0; 1)$.

Посчитаем характеристическую функцию

$$Me^{i\lambda \sum_{j=1}^d \varphi_j \xi_j} = \prod_{j=1}^d Me^{i\lambda \varphi_j \xi_j} =$$

$$= \prod_{j=1}^d e^{\frac{-\lambda^2 \varphi_j^2}{2}} = e^{-\frac{1}{2} \lambda^2 (\sum_{j=1}^d \varphi_j^2)}$$

$$(\vec{\xi}, \vec{\varphi}) \sim N(0, \sum_{j=1}^d \varphi_j^2)$$

Свойства:

1) $\forall j = 1, \dots, d$ ξ_j – гауссовская.

Упражнение. Наоборот не так. Привести контрпример.

$$M\vec{\xi} = \vec{a} = (M\xi_1, \dots, M\xi_d)$$

$$M(\vec{\xi}, \vec{\varphi}) = (M\vec{\xi}, \vec{\varphi}) = (\vec{a}, \vec{\varphi})$$

Пусть η, \varkappa – случайные величины.

Ковариация:

$$M(\eta - M\eta)(\varkappa - M\varkappa) = \text{Cov}(\eta, \varkappa)$$

$A = (a_{ij})_{i,j=1}^d$ – матрица ковариации

$$a_{ij} = M(\xi_i - M\xi_i)(\xi_j - M\xi_j)$$

A – симметрична.

$A \geq 0$ – неотрицательно определена. Проверим:

$\vec{\varphi} \in \mathbb{R}^d$ – произвольный

$$\begin{aligned} (A\vec{\varphi}, \vec{\varphi}) &= \sum_{i,j=1}^d a_{ij}\varphi_i\varphi_j = \\ &= \sum_{i,j=1}^d M(\xi_i - M\xi_i)(\xi_j - M\xi_j)\varphi_i\varphi_j = \\ &= M\left(\sum_{j=1}^d \varphi_j(\xi_j - M\xi_j)\right)^2 \geq 0 \end{aligned}$$

A – ковариационная матрица для вектора $\vec{\xi}$.

Характеристическая функция гауссовского случайного вектора

$$\Phi_{\vec{\xi}}(\vec{\lambda}) = Me^{i(\vec{\lambda}, \vec{\xi})}$$

$(\vec{\lambda}, \vec{\xi})$ – гауссовская

$$M(\vec{\lambda}, \vec{\xi}) = (\vec{\lambda}, \vec{a})$$

$$D(\vec{\lambda}, \vec{\xi}) = (A\vec{\lambda}, \vec{\lambda})$$

$$\Phi_{\vec{\xi}}(\vec{\lambda}) = Me^{i(\vec{\lambda}, \vec{\xi})} = e^{i(\vec{a}, \vec{\lambda}) - \frac{1}{2}(A\vec{\lambda}, \vec{\lambda})}.$$

Распределение гауссовского случайного вектора полностью задается его \vec{a} и A .

Действие линейного оператора на гауссовскую случайную величину $\vec{\xi}$ в \mathbb{R}^d . Пусть $B : \mathbb{R}^d \rightarrow \mathbb{R}^n$ – линейный оператор, $\vec{\eta} = B\vec{\xi}$ – гауссовский случайный вектор

$$(\vec{\eta}, \vec{\psi}) = (B\vec{\eta}, \vec{\psi}) = (\vec{\eta}, B^*\vec{\psi})$$

$$\begin{aligned}
M\vec{\eta} &= BM\vec{\xi} = B\vec{a} \\
D(\vec{\eta}, \vec{\psi}) &= D(\vec{\xi}, B^*\vec{\psi}) = \\
&= (AB^*\vec{\psi}, B^*\vec{\psi}) = (BAB^*\vec{\psi}, \vec{\psi}).
\end{aligned}$$

Таким образом, если $\vec{\xi} \rightarrow B\vec{\xi}$, то $\vec{a} \rightarrow B\vec{a}$, $A \rightarrow BAB^*$.

Лемма. $\vec{a} \in \mathbb{R}^d$; $A = A^*$; $A \geq 0$. Тогда существует гауссовский случайный вектор с параметрами \vec{a} и A .

▷ Достаточно доказать для случая $\vec{a} = \vec{0}$.

В некотором ортонормированном базисе

$$A = \begin{pmatrix} \lambda_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_n \end{pmatrix}$$

$$\forall j = 1, \dots, d; \lambda_j \geq 0$$

$$B = \begin{pmatrix} \sqrt{\lambda_1} & \dots & 0 \\ & \ddots & \\ 0 & \dots & \sqrt{\lambda_d} \end{pmatrix}$$

$$B = B^*, B^2 = A$$

Пусть $\vec{\xi}$ – стандартный гауссовский случайный вектор в \mathbb{R}^d
 $\vec{\eta} = B\vec{\xi}$

$$M\vec{\eta} = 0; A_{\vec{\eta}} = BA_{\vec{\xi}}B^*$$

Упражнение. $A_{\vec{\xi}} = I$

$$A_{\vec{\eta}} = BIB^* = BB^* = B^2 = A. \quad \triangleleft$$

Лемма. Гауссовский случайный вектор с параметрами \vec{a} , A имеет плотность тогда и только тогда, когда $A > 0$ ($\det A \neq 0$). В этом случае плотность имеет вид

$$p(\vec{x}) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det A}} e^{-\frac{1}{2}(A^{-1}(\vec{x}-\vec{a}), \vec{x}-\vec{a})}.$$

▷ Пусть $\vec{\xi}$ – гауссовский стандартный вектор в \mathbb{R}^d . Он имеет плотность, т.к. все координаты имеют плотность и независимы:

$$p_{\vec{\xi}}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\|\vec{x}\|^2}$$

Пусть $\vec{\eta} = B\vec{\xi}$, где B – как в предыдущей лемме.

1) $\det B = 0 \Leftrightarrow \det A = 0$

$B(\mathbb{R}^d) = L$ – собственное подпространство \mathbb{R}^d .

Мера Лебега для L равна 0.

$$P(\vec{\eta} \in L) = 1,$$

отсюда следует, что $\vec{\eta}$ не имеет плотности.

2) $\det B > 0 \Leftrightarrow \det A > 0$

$$\begin{aligned} Mf(\vec{\eta}) &= Mf(B\vec{\xi}) = \int_{\mathbb{R}^d} f(B\vec{y}) \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{\|\vec{y}\|^2}{2}} d\vec{y} = \\ &= \left| \begin{array}{l} B\vec{y} = \vec{x} \\ \vec{y} = B^{-1}\vec{x} \\ d\vec{y} = \det B^{-1} d\vec{x} = \frac{1}{\sqrt{\det A}} d\vec{x} \end{array} \right| = \int_{\mathbb{R}^d} f(\vec{x}) \frac{1}{(\sqrt{2\pi})^d \sqrt{\det A}} e^{-\frac{1}{2}\|B^{-1}\vec{x}\|^2} d\vec{x} \\ \|B^{-1}\vec{x}\|^2 &= (B^{-1}\vec{x}, B^{-1}\vec{x}) = (B^{-2}\vec{x}, \vec{x}) = (A^{-1}\vec{x}, \vec{x}) \end{aligned}$$

$\forall \vec{a}$ сдвигаем $\vec{\eta}$ на \vec{a} . \triangleleft

Пример. $\vec{\xi}$ – гауссовский случайный вектор с параметрами $\vec{0}$, A , $A > 0$.

Рассмотрим семейство

$$\{\vec{\theta} + \vec{\xi}, \vec{\theta} \in \mathbb{R}^d\}$$

Пусть $\vec{\varphi} \in \mathbb{R}^d$. Какова наилучшая несмещенная оценка для $(\vec{\varphi}, \vec{\theta})$?
 $\vec{\psi} \in \mathbb{R}^d$

$$U_{\vec{\psi}}(\vec{\theta}, \vec{x}) = \frac{\partial}{\partial_{\vec{\psi}} \vec{\theta}} \ln(\vec{\theta}, \vec{x}) = (A^{-1}(\vec{x} - \vec{\theta}), \vec{\psi})$$

$$\begin{aligned} I_{\vec{\psi}}(\vec{\theta}) &= M(A^{-1}\vec{\xi}, \vec{\psi})^2 = M(\vec{\xi}, A^{-1}\vec{\psi})^2 = \\ &= (AA^{-1}\vec{\psi}, A^{-1}\vec{\psi}) = (A^{-1}\vec{\psi}, \vec{\psi}). \end{aligned}$$

α – несмещенная оценка для $(\vec{\varphi}, \vec{\theta})$:

$$\forall \vec{\psi} \neq \vec{0} \quad D\alpha \geq \frac{(\vec{\varphi}, \vec{\psi})^2}{(A^{-1}\vec{\psi}, \vec{\psi})}$$

$$D\alpha \geq \sup_{\vec{\psi} \neq \vec{0}} \frac{(\vec{\varphi}, \vec{\psi})}{(A^{-1}\vec{\psi}, \vec{\psi})}$$

$$A^{-\frac{1}{2}}\vec{\psi} = \vec{\gamma} \in \mathbb{R}^d \setminus \{\vec{0}\}$$

$$D\alpha \geq \sup_{\vec{\gamma}=\vec{0}} \frac{A^{1/2}(\vec{\varphi}, \vec{\gamma})^2}{(\vec{\gamma}, \vec{\gamma})} = \|A^{\frac{1}{2}}\vec{\varphi}\|^2 = (A\vec{\varphi}, \vec{\varphi}).$$

$$D\alpha \geq (A\vec{\varphi}, \vec{\varphi})$$

$$\alpha^* = (\vec{\varphi}, \vec{x})$$

$$M\alpha^* = (\vec{\varphi}, \vec{\theta})$$

$$D\alpha^* = (A\vec{\varphi}, \vec{\varphi})$$

$\alpha^* = (\vec{\varphi}, \vec{x})$ – эффективная несмещенная оценка

Лекция 8

Теорема о нормальной корреляции.

Квадратичные формы от гауссовских случайных векторов

$$\mathbb{R}^d; \vec{\xi} = (\xi_1, \dots, \xi_d)$$

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = e^{i(\vec{a}, \vec{\lambda}) - \frac{1}{2}(A\vec{\lambda}, \vec{\lambda})}$$

$$\vec{a} = M\vec{\xi}$$

$$A = \|M(\xi_i - M\xi_i)(\xi_j - M\xi_j)\|_{i,j=1}^d = \text{Cov}_{\vec{\xi}, \vec{\xi}}$$

$\vec{\xi}, \vec{\eta}$ – гауссовские случайные вектора такие, что $(\xi_1, \dots, \xi_d, \eta_1, \dots, \eta_n)$ – гауссовский случайный вектор.

Лемма. $\vec{\xi}$ и $\vec{\eta}$ – независимы $\Leftrightarrow \text{Cov}_{\vec{\xi}, \vec{\eta}} = 0$

▷

$$(\xi_1, \dots, \xi_d, \eta_1, \dots, \eta_n) =: \vec{\mathcal{X}}$$

$$\varphi_{\vec{\mathcal{X}}}(\vec{\lambda}) = e^{i(\vec{a}, \vec{\lambda}) - \frac{1}{2}(A\vec{\lambda}, \vec{\lambda})}$$

$$\vec{a} = (M\xi_1, \dots, M\xi_d, M\eta_1, \dots, M\eta_n)$$

$$A = \left(\begin{array}{c|c} \text{Cov}_{\vec{\xi}, \vec{\xi}} & \text{Cov}_{\vec{\xi}, \vec{\eta}} \\ \hline \text{Cov}_{\vec{\eta}, \vec{\xi}} & \text{Cov}_{\vec{\eta}, \vec{\eta}} \end{array} \right)$$

$$\vec{\lambda} = (u_1, \dots, u_d, v_1, \dots, v_n)$$

$$\vec{u} = (u_1, \dots, u_d)$$

$$\vec{v} = (v_1, \dots, v_n)$$

$$\vec{\xi} \text{ и } \vec{\eta} \text{ – независимы } \Leftrightarrow \varphi(\vec{\lambda}) = \varphi_{\vec{\xi}}(\vec{u})\varphi_{\vec{\eta}}(\vec{v}) \quad (\forall \vec{u}, \vec{v}).$$

$$\text{Cov}_{\vec{\xi}, \vec{\eta}} = 0.$$

◁

Упражнение. Пример случайных величин, являющихся некоррелированными, но зависимыми.

(Ω, \mathcal{F}, P) ; $\mathcal{F}' \subset \mathcal{F}$ – под- σ -алгебра

УМО: $M(\eta/\mathcal{F}') = \zeta$:

1) ζ – измерима относительно \mathcal{F}'

2) $\forall \Delta \in \mathcal{F}' : M\eta \mathbb{I}_\Delta = M\zeta \mathbb{I}_\Delta$

$M(\cdot/\mathcal{F}')$ – проектор в L_2 на подпространство случайных величин, измеримых относительно \mathcal{F}' .

Если $\mathcal{F}' = \sigma(\xi)$, то

$$M(\eta/\sigma(\xi)) = \varphi(\xi).$$

Лемма. Пусть случайная величина \varkappa не зависит от случайной величины ξ . Тогда

$$M(\xi + \varkappa/\xi) = \xi + M\varkappa.$$

▷

$$M(\xi/\xi) = \xi$$

$$M(\varkappa/\xi) = M\varkappa$$

Пусть $\vec{\xi}, \vec{\eta}$ – совместно гауссовские

$$M(\vec{\eta}/\vec{\xi}) = ?$$

Пример. $\vec{\xi}, \vec{\eta}$ – одномерные

$$M\xi = M\eta = 0$$

$$\varkappa = \eta - \left(\frac{M\xi\eta}{M\xi^2} \right) \xi$$

$$M(\varkappa\xi) = 0$$

\varkappa и ξ – некоррел. \Rightarrow независимы.

$$\eta = \left(\frac{M\xi\eta}{M\xi^2} \right) \xi + \varkappa$$

$$M(\eta/\xi) = \left(\frac{M\xi\eta}{M\xi^2} \right) \xi$$

Условное распределение случайной величины η – это нормальное распределение со средним $\frac{M\xi\eta}{M\xi^2} M\xi$ и дисперсией:

$$\begin{aligned} M\varkappa^2 &= M\eta^2 - 2 \frac{(M\xi\eta)^2}{M\xi^2} + \frac{(M\xi\eta)^2}{M\xi^2} = \\ &= M\eta^2 - \frac{(M\xi\eta)^2}{M\xi^2} \end{aligned}$$

$$M(\vec{\eta}/\vec{\xi}) = (M(\eta_1/\xi), \dots, M(\eta_n/\xi))$$

Теорема (о нормальной корреляции). Пусть $\vec{\xi}, \vec{\eta}$ – совместно гауссовские; $M\vec{\xi} = \vec{0}$, $M\vec{\eta} = \vec{0}$, $\text{Cov}_{\vec{\xi}, \vec{\eta}} > 0$. Тогда а) $M(\vec{\eta}/\vec{\xi}) = \text{Cov}_{\vec{\eta}, \vec{\xi}} \text{Cov}_{\vec{\xi}, \vec{\xi}}^{-1} \vec{\xi}$.

б) Условное распределение $\vec{\eta}$ при известном $\vec{\xi}$ – гауссовское со средним из п. а) и ковариационной матрицей:

$$\text{Cov}_{\vec{\eta}, \vec{\eta}} - \text{Cov}_{\vec{\eta}, \vec{\xi}} \text{Cov}_{\vec{\xi}, \vec{\xi}}^{-1} \text{Cov}_{\vec{\xi}, \vec{\eta}}.$$

▷ ◁

Следствия. Пусть $\vec{\xi}$ – стандартный гауссовский случайный вектор в \mathbb{R}^d , тогда

$$\vec{a} = \vec{0}, \quad \text{Cov}_{\vec{\xi}, \vec{\xi}} = I.$$

Пусть $\vec{e}, \vec{g} \in \mathbb{R}^d$; $\vec{e} \perp \vec{g}$, тогда

$(\vec{e}, \vec{\xi}), (\vec{g}, \vec{\xi})$ – независимы.

▷ $(\vec{e}, \vec{\xi}), (\vec{g}, \vec{\xi})$ – совм. гауссовские

$\vec{\xi}, \vec{g}$ – имеют нулевые средние.

$$M(\vec{e}, \vec{\xi})(\vec{g}, \vec{\xi}) = (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{e}, \vec{g}) = (\vec{e}, \vec{g}) = 0. \quad \triangleleft$$

Пусть $\eta = (A\vec{\xi}, \vec{\xi})$, $A \geq 0$.

Ортонормированный базис $\{\vec{e}_1, \dots, \vec{e}_d\}$ существует, причем в нем $A =$

$$\begin{pmatrix} \lambda_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_d \end{pmatrix}$$

$$\eta = \sum_{k=1}^d \lambda_k (\vec{e}_k, \vec{\xi})^2, \quad (\vec{e}_k, \vec{\xi}) = \varkappa_k$$

$(\varkappa_1, \dots, \varkappa_d)$ – независимые одинаково распределенные, имеющие стандартное гауссовское распределение $N(0; 1)$

$$Me^{it\eta} = Me^{it \sum_{k=1}^d \lambda_k \varkappa_k^2} = \prod_{k=1}^d Me^{it \lambda_k \varkappa_k^2}$$

Лекция 9

Наблюдается вектор \vec{Y} : $\vec{Y} = A\vec{\theta} + \vec{\xi}$, $\vec{\xi}$ случайный вектор

$A : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $\vec{\theta}$ – неизвестные параметры, $\vec{\xi}$ – случайный вектор с $M\vec{\xi} = \vec{0}$.

$\vec{\theta} = ?$

Предположения:

$$m \leq n; \text{rang } A = m; \text{Cov}_{\vec{\xi}, \vec{\xi}} = I$$

$$\vec{\xi} \sim \text{Gauss}(\vec{0}, \text{Cov}_{\vec{\xi}, \vec{\xi}})$$

Найдем о.м.п. для $\vec{\theta}$

$$L(\vec{y}, \vec{\theta}) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(\vec{y} - A\vec{\theta}, \vec{y} - A\vec{\theta})}$$

$$\max_{\vec{\theta}} L(\vec{y}, \vec{\theta}) \Leftrightarrow \min_{\vec{\theta}} \{\|\vec{y} - A\vec{\theta}\|\} \quad (*)$$

Рис.

$$\forall \vec{h} \in \mathbb{R}^m$$

$$(\vec{y} - A\vec{\theta}_*, A\vec{h}) = 0$$

$$A^* \vec{y} - A^* A \vec{\theta}_* = \vec{0}$$

$A^* A$ – невырожденный ($\text{rang } A = m$), обратимый (положительный) оператор в \mathbb{R}^m .

Следовательно,

$$\vec{\theta}_* = (A^* A)^{-1} A^* \vec{Y}$$

О.м.п. $\vec{\theta}_*$ – линейная, следовательно является оптимальной для любого (даже негауссовского) вектора $\vec{\xi}$ с единичной ковариацией.

Так как ответ получается решением минимизационной задачи (*), полученная оценка называется оценкой метода наименьших квадратов.

Теорема (Гаусса–Маркова). Пусть $L\vec{Y}$ – линейная оценка для $T\vec{\theta}$, где $T : \mathbb{R}^m \rightarrow \mathbb{R}^k$, являющаяся несмещенной. Тогда $\forall \vec{\varphi} \in \mathbb{R}^k$

$$D_{\vec{\theta}}(L\vec{Y}, \vec{\varphi}) \geq D_{\vec{\theta}}(T\vec{\theta}_*, \vec{\varphi}),$$

где $\vec{\theta}_*$ – оценка метода наименьших квадратов (наилучшей линейной оценкой функции от параметра, есть эта же функция от оценки м.н.к.)

▷ в гауссовском случае.

Пусть $\vec{\xi}$ – стандартный гауссовский вектор.

Запишем неравенство Рао–Крамера. Пусть \vec{x} – несмещенная оценка для $T\vec{\theta}$.

$\vec{\varphi} \in \mathbb{R}^k$ – произвольный, тогда $(\vec{x}, \vec{\varphi})$ – несмещенная оценка для $(T\vec{\theta}, \vec{\varphi})$

$$L(\vec{Y}, \vec{\theta}) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\|\vec{Y} - A\vec{\theta}\|^2}$$

$\vec{\psi} \in \mathbb{R}^m$ – произвольный

$$\partial_{\vec{\psi}} \ln L(\vec{Y}, \vec{\theta}) = (\vec{Y} - A\vec{\theta}, A\vec{\psi})$$

$$M(\partial_{\vec{\psi}} \ln L(\vec{Y}, \vec{\theta}))^2 = M(\vec{\xi}, A\vec{\psi})^2 =$$

$$= (A\vec{\psi}, A\vec{\psi}) = (A^* A\vec{\psi}, \vec{\psi})$$

$$D_{\vec{\theta}}(\vec{\mathcal{Z}}, \vec{\varphi}) \geq \frac{|\partial_{\vec{\psi}}(T\vec{\theta}, \vec{\varphi})|}{(A^* A\vec{\psi}, \vec{\psi})}$$

$$D_{\vec{\theta}}(\vec{\mathcal{Z}}, \vec{\varphi}) \geq \sup_{\vec{\psi} \neq \vec{0}} \frac{|(T\vec{\psi}, \vec{\varphi})|}{(A^* A\vec{\psi}, \vec{\psi})} =$$

$$= \sup_{\vec{\psi} \neq \vec{0}} \frac{|(\vec{\psi}, T^* \vec{\varphi})|}{(A^* A\vec{\psi}, \vec{\psi})} =$$

$$= ((A^* A)^{-1} T^* \vec{\varphi}, T^* \vec{\varphi})$$

$$\vec{\theta}_* = (A^* A)^{-1} A^* \vec{Y}$$

$$D_{\vec{\theta}}(T\vec{\theta}_*, \vec{\varphi}) = D_{\vec{\theta}}(\vec{\theta}_*, T^* \vec{\varphi})$$

$$M\vec{\theta}_*, \text{Cov}_{\vec{\theta}_*, \vec{\theta}_*} - ?$$

$$\vec{\theta}_* = (A^* A)^{-1} A^* (A\vec{\theta} + \vec{\xi}) = \vec{\theta} + (A^* A)^{-1} A^* \vec{\xi}$$

$$M\vec{\theta}_* = \vec{\theta}$$

$$\text{Cov}_{\vec{\theta}_*, \vec{\theta}_*} = (A^* A)^{-1} A^* A (A^* A)^{-1} = (A^* A)^{-1}$$

$$D_{\vec{\theta}}(T\vec{\theta}_*, \vec{\varphi}) = (\text{Cov}_{\vec{\theta}_*, \vec{\theta}_*} T^* \vec{\varphi}, T^* \vec{\varphi}). \quad \triangleleft$$

Как общий случай привести к виду

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} = I \quad ?$$

$$\vec{Y} = A\vec{\theta} + \vec{\xi}, \quad m \leq n, \quad \text{rank } A = m,$$

$$M\vec{\xi} = \vec{0}, \quad \text{Cov}_{\vec{\xi}, \vec{\xi}} = V > 0, \quad V \neq I.$$

$$\underbrace{V^{-\frac{1}{2}} \vec{Y}}_{\vec{Z}} = \underbrace{V^{-\frac{1}{2}} A}_{\tilde{A}} \vec{\theta} + \underbrace{V^{-\frac{1}{2}} \vec{\xi}}_{\vec{\eta}}$$

$$\vec{Z} = \tilde{A} \vec{\theta} + \vec{\eta}$$

$$M\vec{\eta} = \vec{0}$$

$$\text{Cov}_{\vec{\eta}, \vec{\eta}} = V^{-\frac{1}{2}} V V^{-\frac{1}{2}} = I$$

Получается предыдущая схема.

Надо проверить, чему равен $\text{rank } V^{-\frac{1}{2}} A$

$$\text{rank } V^{-\frac{1}{2}} A = m$$

$$\vec{\theta}_* = (\tilde{A}^* \tilde{A})^{-1} \tilde{A}^* \vec{Z}$$

Пример. Линейная регрессия.

$x, y \in \mathbb{R}$

$y = ax + b$ $y_k = ax_k + b + \xi_k$, – наблюдения $\vec{\xi} \sim N(0; I)$

$a := \theta_1$; $b := \theta_2$

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}; \quad \text{rank } A = 2$$

$$\vec{\theta}_* = (A^* A)^{-1} A^* \vec{Y}$$

$$A^* = \begin{pmatrix} x_1 & \dots & x_n \\ 1 & \dots & 1 \end{pmatrix}$$

$$A^* A = \begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix}$$

$$\det A^* A = n \sum x_k^2 - (\sum x_k)^2$$

$$(A^* A)^{-1} = \frac{1}{n \sum x_k^2 - (\sum x_k)^2} \begin{pmatrix} n & -\sum_{i=1}^n x_k \\ -\sum_{i=1}^n x_k & \sum x_k^2 \end{pmatrix}$$

$$A^* \vec{Y} = \begin{pmatrix} \sum x_k y_k \\ \sum y_k \end{pmatrix}$$

$$\vec{\theta}_* = \frac{1}{n \sum x_k^2 - (\sum x_k)^2} \begin{pmatrix} n \sum x_k y_k - \sum x_k \sum y_k \\ \sum x_k^2 \sum y_k - \sum x_k \sum x_k y_k \end{pmatrix}$$

Пусть $\hat{\theta}$ – эффективная оценка θ

$$D_{\theta} \hat{\theta} = \frac{1}{I_n(\theta)}$$

Иногда $\hat{\theta}$ слишком сильно отличается от θ .

Рис.

Доверительные интервалы

По выборке находим два числа

$$T_1 = T_1(x_1, \dots, x_n); T_2 = T_2(x_1, \dots, x_n); \alpha \in (0; 1) \text{ – уровень доверия}$$

Определение. $[T_1; T_2]$ – доверительный интервал для названного значения параметра θ с уровнем доверия α , если

$$P_{\theta}\{\tau_1 \leq \theta \leq \tau_2\} \geq 1 - \alpha; \quad \forall \theta \in \Theta$$

Рис.

Пример.

1. (с помощью неравенства Чебышева)

$$N(\theta; 1); \theta \in \mathbb{R}$$

$$\bar{x} \sim N(\theta; \frac{1}{n})$$

$$P\{|\bar{x} - \theta| > \varepsilon\} \leq \frac{1}{n\varepsilon^2}$$

$$P\{\theta \in [\bar{x} - \varepsilon; \bar{x} + \varepsilon]\} \geq 1 - \frac{1}{n\varepsilon^2}$$

ε выбрано так, что $\frac{1}{n\varepsilon^2} \leq \alpha$. Тогда $[\bar{x} - \varepsilon; \bar{x} + \varepsilon]$ – доверительный интервал.
 $T_1 = \bar{x} - \varepsilon; T_2 = \bar{x} + \varepsilon$.

$$l = T_2 - T_1 = 2\varepsilon \geq \frac{2}{\sqrt{n\alpha}}.$$

Упражнение. Построить доверительные интервалы, пользуясь обобщенным неравенством Чебышева для функции x^m

$$\begin{aligned} P\{\theta \in [\bar{x} - \varepsilon; \bar{x} + \varepsilon]\} &= \\ &= P\{\bar{x} \in [\theta - \varepsilon; \theta + \varepsilon]\} = \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{\theta - \varepsilon}^{\theta + \varepsilon} e^{-\frac{n(u - \theta)^2}{2}} du = \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{nv^2}{2}} dv \leq 1 - \alpha. \end{aligned}$$

Определение. Функция $G(x_1, \dots, x_n; \theta)$ называется центральной статистикой, если

- 1) распределение $G(\vec{x}, \theta)$ не зависит от θ ,
- 2) $G(x, \theta)$ – строго монотонна и непрерывна по θ .

В примере 1

$$G(\vec{x}, \theta) = \bar{x} - \theta \sim N(0; \frac{1}{n})$$

и линейная функция по θ .

Пусть $G(\vec{x}, \theta)$ – центральная статистика. $\alpha \mapsto g_1, g_2 \in \mathbb{R}$ ($g_1 < g_2$) :

$$P\{G(\vec{x}, \theta) \in [g_1, g_2]\} \geq 1 - \alpha$$

$$g_1 \leq G(\vec{x}, \theta) \leq g_2$$

$G(\vec{x}, \theta)$ – строго монотонна и непрерывна по θ , получаем неравенство

$$T_1(x) \leq \theta \leq T_2(x)$$

$$P\{\theta \in [T_1; T_2]\} \geq 1 - \alpha$$

В примере 1 отрезок $[g_1, g_2]$ выбирается неоднозначно.

Рис.

Надо брать так, чтобы $g_2 - g_1$ была минимальной.

Упражнение. Наименьшая длина интервала, когда $g_1 = -g_2$.

Как найти центральную статистику?

Лемма. ξ – случайная величина, имеющая непрерывную строго возрастающую функцию распределения F . Тогда $F(\xi)$ равномерно распределена на $[0; 1]$.

$$\triangleright C \in [0; 1]$$

$$\begin{aligned} P\{F(\xi) \leq C\} &= P\{\xi \leq F^{-1}(C)\} = \\ &= F(F^{-1}(C)) = C. \quad \triangleleft \end{aligned}$$

$F_\theta, \theta \in \Theta$. Пусть $\forall \theta$ F_θ строго монотонно и непрерывна

$$G(x, \theta) = \sum_{k=1}^n \ln F_\theta(x_k)$$

$G(x, \theta)$ имеет известное распределение (Эрланга).

x_1, \dots, x_n – выборка, θ – неизвестное среднее. Дисперсия известна σ^2

$$\frac{\sqrt{n}(\bar{x} - \theta)}{\sigma} \sim N(0; 1) \quad (\text{ЦПТ})$$

Лекция 10

В предыдущих схемах распределения выборки или зависимость наблюдения от параметра точки предполагалась известной. Как определить является ли используемая схема адекватной?

Проверка статистических гипотез

1. Является ли набор случайных величин (x_1, \dots, x_n) выборкой из распределения F ?

2. Являются ли случайные величины (x_1, \dots, x_n) независимы?

3. Пусть известно, что (x_1, \dots, x_n) и (y_1, \dots, y_n) – выборки из распределений $F_\theta, \theta \in \Theta$. Совпадают ли параметры θ_x и θ_y ?

4. Является ли набор (x_1, \dots, x_n) выборкой из распределения F_1 или он является выборкой из распределения F_2 ?

Критерии согласия

Ответ на первый вопрос: $D \subset \mathbb{R}^n$.

Если $\vec{x} \in D$ – да ($\vec{x} = (x_1, \dots, x_n)$)

$\vec{x} \notin D$ – нет.

$P_F(\vec{x} \notin D)$ может быть положительной. P_F – вероятность когда \vec{x} из распределения F .

$\alpha \in (0; 1)$ и выбираем $D \subset \mathbb{R}^n$:

$$P_F(\vec{x} \notin D) \leq \alpha.$$

В вопросе 4: вероятность ошибки $P_{F_2}(\vec{x} \in D) \searrow$. Сейчас вероятности ошибки нет. Поэтому надо искать $D : \sup_{F'} P_{F'}(\vec{x} \in D) \searrow$.

Надо найти D :

$$P_F(\vec{x} \notin D) \leq \alpha, \forall F' \neq F \quad P_{F'}(\vec{x} \notin D) \geq \alpha$$

D – несмещенный критерий согласия.

Будем искать D как линии уровня некоторой функции:

$$\Phi : \mathbb{R}^n \rightarrow \mathbb{R}, D = \{\Phi \leq C\}$$

$\Phi(x_1, \dots, x_n)$ – статистика. Если знаем распределение Φ , можно считать $P_F\{\Phi \leq C\}$.

Пусть Φ имеет известное распределение, если выборка сделана из распределения F .

Можно построить критерий:

$$\alpha \rightarrow C_\alpha : P_F\{\Phi \leq C_\alpha\} = \alpha$$

$D_\alpha = \{\Phi \leq C_\alpha\}$ – искомая крит. области.

Надо найти Φ с хорошим распределением, и чтобы распределение сильно менялось при замене F на $F' \neq F$.

Критерий Колмогорова

F – непрерывная и строго монотонная

$$\Phi(\vec{x}) = \sup_{u \in \mathbb{R}} |F_n(u) - F(u)|,$$

где $F_n(u)$ – эмпирическая функция распределения.

$$F_n(u) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{[x_k; +\infty)}(u)$$

Лемма. Φ имеет известное распределение, если \vec{X} – выборка из F .

▷

$$\begin{aligned} \sup_{\mathbb{R}} |F_n(u) - F(u)| &= \sup_{[0;1]} |F_n(F^{-1}(v)) - F(F^{-1}(v))| = \\ &= \sup_{[0;1]} |v - F_n(F^{-1}(v))| = \\ &= \sup_{[0;1]} \left| v - \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{[F^{-1}(F(x_k)); +\infty)}(F^{-1}(v)) \right| = \\ &= \sup_{[0;1]} \left| v - \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{[(F(x_k)); +\infty)}(v) \right|, \end{aligned}$$

где $\frac{1}{n} \sum_{k=1}^n \mathbb{I}_{[(F(x_k)); +\infty)}(v)$ – эмпирическая функция, построенная по выборке $(F(x_1), \dots, F(x_n))$ из равномерного распределения.

Т.о. распределение Φ не зависит от F . ◁

При $n \rightarrow \infty$ Φ , записанное для другого распределения, не может быть сколь угодно маленькой.

Критерий χ^2

F ;

Рис.

$$P_F(x \in \Delta_k) = p_k, \quad p_1, \dots, p_N > 0, \quad \sum_{k=1}^n p_k = 1$$

$$x_1, \dots, x_n \mapsto \nu_k = \sum_{j=1}^n \mathbb{I}_{\Delta_k}(x_j)$$

$$\Phi_n = \sum_{k=1}^N \frac{(\nu_k - np_k)^2}{np_k}$$

Теорема. $\Phi_n \Rightarrow \chi_{N-1}^2, n \rightarrow \infty$

(χ_{N-1}^2 – распределение $\sum_{s=1}^{N-1} \xi_s^2, \{\xi_s\}$ – независимые, $N(0; 1)$)

▷

$$\eta_n = \left(\frac{\nu_1 - np_1}{\sqrt{n}}, \frac{\nu_2 - np_2}{\sqrt{n}}, \dots, \frac{\nu_N - np_N}{\sqrt{n}} \right)$$

– случайный вектор в \mathbb{R}^N .

$$Me^{i(\lambda, \eta_n)} = Me^{i \sum_{k=1}^N \lambda_k \frac{\nu_k - np_k}{\sqrt{n}}} = e^{-i \sum_{k=1}^N \sqrt{n} \lambda_k p_k} \times$$

$$\times Me^{i \sum_{k=1}^N \frac{\lambda_k}{\sqrt{n}} \sum_{j=1}^n \mathbb{I}_{\Delta_k}(x_j)} = e^{-i \sum_{k=1}^N \lambda_k p_k \sqrt{n}} \times$$

$$\begin{aligned}
& \times \left(M e^{i \sum_{k=1}^N \frac{\lambda_k}{\sqrt{n}} \sum_{j=1}^n \mathbb{I}_{\Delta_k}(x_j)} \right)^n = \\
& = e^{-i \sum_{k=1}^N \lambda_k p_k \sqrt{n}} \left(\sum_{k=1}^N e^{i \frac{\lambda_k}{\sqrt{n}} p_k} \right)^n = \\
& = e^{-i \sum_{k=1}^N \lambda_k p_k \sqrt{n}} \left(1 + \sum_{k=1}^N \left(e^{i \frac{\lambda_k}{\sqrt{n}}} - 1 \right) p_k \right)^n \quad (*) \\
& \ln(1+u) \sim u - \frac{u^2}{2} + O(u^3), u \rightarrow 0 \\
& e^v - 1 \sim v + \frac{v^2}{2} + O(v^3), v \rightarrow 0
\end{aligned}$$

Логарифмируем (*)

$$\begin{aligned}
& -i \sum_{k=1}^N \lambda_k p_k \sqrt{n} + n \ln \left(1 + \sum_{k=1}^N \left(e^{i \frac{\lambda_k}{\sqrt{n}}} - 1 \right) p_k \right) = \\
& = -i \sum_{k=1}^N \lambda_k p_k \sqrt{n} + n \sum_{k=1}^N \left(e^{i \frac{\lambda_k}{\sqrt{n}}} - 1 \right) p_k - \\
& - \frac{n}{2} \left(\sum_{k=1}^N \left(e^{i \frac{\lambda_k}{\sqrt{n}}} - 1 \right) p_k \right)^2 = -i \sum_{k=1}^N \lambda_k p_k \sqrt{n} + \\
& + n \sum_{k=1}^N i \frac{\lambda_k}{\sqrt{n}} p_k + n \sum_{k=1}^n \frac{p_k}{2} \left(-\frac{\lambda_k^2}{n} \right) - \dots = \\
& = -\frac{1}{2} \sum_{k=1}^n p_k \lambda_k^2 \\
& (*) \xrightarrow{n \rightarrow \infty} -\frac{1}{2} (S\lambda, \lambda)
\end{aligned}$$

где $S = (\delta_{k_1 k_2} p_{k_1 k_2} - p_{k_1} p_{k_2})$, $S \geq 0$

$$M e^{i(\lambda, \eta_n)} \rightarrow e^{-\frac{1}{2}(S\lambda, \lambda)}, \quad \eta_n \Rightarrow N(0, S)$$

Пусть ζ – гауссовский случайный вектор с распределением $N(0, S)$

$$\Phi_n = (A\eta_n, A\eta_n),$$

где

$$A = \begin{pmatrix} \frac{1}{\sqrt{p_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{p_N}} \end{pmatrix}$$

Φ_n – непрерывная функция от $\eta_n, \eta_n \Rightarrow \zeta$

$$\Phi_n \Rightarrow (A\zeta, A\zeta), n \rightarrow \infty$$

$\zeta = S^{1/2}\xi$, где ξ – стандартный гауссовский вектор в \mathbb{R}^N

$$(A\zeta, A\zeta) = (AS^{1/2}\xi, AS^{1/2}\xi) = (S^{1/2}A^2S^{1/2}\xi, \xi)$$

$$\text{Cov}_{A\zeta, A\zeta} = A^* \text{Cov}_{\zeta, \zeta} A = ASA$$

$(A\zeta, A\zeta)$ имеет такое же распределение, как и $(ASA\xi', \xi')$, ξ' – стандартный гауссовский в \mathbb{R}^N .

$$\varkappa \stackrel{d}{=} N(0; V)$$

$$\varkappa = V^{1/2}\xi', \quad (\varkappa, \varkappa) = (V\xi', \xi')$$

$$ASA = (\delta_{k_1 k_2} - \sqrt{p_{k_1}} \sqrt{p_{k_2}})$$

$$(Q\xi, \xi) \stackrel{d}{=} \sum_{j=1}^N \lambda_j (\xi'_j)^2$$

λ_j – собственные числа Q , $\{\xi'_j\}$ – независимые, $N(0; 1)$

1. $\|e\| = 1, \quad e = (e_1, \dots, e_N)$

$P_e(x) = (e, x)e$ – проектор на e

$\Pi = (e_{k_1} e_{k_2})_{k_1 k_2=1}^N$ – квадратичная форма от e

$I - \Pi$ в базисе из собственных векторов имеет вид:

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

◁

Лекция 11

Критерий Пирсона (χ^2) для проверки независимости двух величин

Наблюдение двух случайных величин $x_1, \dots, x_n, y_1, \dots, y_n$.

1) Области значений для $\{x_n\}, \{y_n\}$ делятся на интервалы

$I_1, \dots, I_k \rightarrow x,$

$$J_1, \dots, J_l \rightarrow y.$$

Рис.

ν_{ij} равно количеству испытаний, в которых: $x \in I_i, y \in J_j$

p_i — частота, с которой x попадает в I_i

q_j — частота, с которой y попадает в J_j

$$\sum_{i,j} \frac{(\nu_{ij} - np_i q_j)^2}{np_i q_j} =: T_0$$

Произведение появляется с учетом независимости случайных величин.

Утверждение. $T_0 \Rightarrow \chi_{kl-(k+l-1)}^2 = \chi_{(k-1)(l-1)}^2, n \rightarrow \infty.$

▷ ◁

В T_0 значения p_i и q_j могут быть известны лишь при некоторых априорных предположениях. Поэтому p_i и q_j заменены их выборочными аналогами.

$$\begin{aligned} p_i &\rightarrow \frac{1}{n} \sum_{r=1}^n \mathbb{I}_{\{x_r \in I_i\}} \\ \nu_{i,\cdot} &= \sum_{j=1}^l \nu_{ij}, \quad \nu_{\cdot,j} = \sum_{i=1}^l \nu_{ij}. \\ p_i &\rightarrow \frac{1}{n} \nu_{i,\cdot}, \quad q_j \rightarrow \frac{1}{n} \nu_{\cdot,j}. \\ T &= \sum_{i,j} \frac{(\nu_{ij} - n \frac{\nu_{i,\cdot}}{n} \frac{\nu_{\cdot,j}}{n})^2}{n \frac{\nu_{i,\cdot}}{n} \frac{\nu_{\cdot,j}}{n}} = \\ &= \sum_{\substack{i=1,\overline{k} \\ j=1,\overline{l}}} \frac{n(\nu_{ij} - \frac{\nu_{i,\cdot} \nu_{\cdot,j}}{n})^2}{\nu_{i,\cdot} \nu_{\cdot,j}}. \end{aligned}$$

Упражнение. $f \in C(\mathbb{R}^2)$

$$\xi_n \Rightarrow \xi, \eta_n \xrightarrow{P} C = \text{const} \Rightarrow S_n = f(\xi_n, \eta_n) \Rightarrow f(\xi, C)$$

(тогда $T \Rightarrow \chi_{(k-1)(l-1)}^2$).

Схема проверки критерия.

α — уровень доверия.

1. Разбиваем область значения на интервалы (дискретизация).

2) Считаем значение статистики T .

3) Находим $\gamma_{\alpha, (k-1)(l-1)}$ – квантиль для распределения $\chi^2_{(k-1)(l-1)}$ уровня α и проверяем, выполняется ли соотношение $T < \gamma_{\alpha, (k-1)(l-1)}$.

Да \rightarrow принимаем

Нет \rightarrow отвергаем.

Квантиль уровня $\frac{1}{2}$ называется медианой.

Если $z_{11} = P(x \in I_1, y \in J_1) \neq p_1 q_1$

$$T_0 = \frac{(\nu_{11} - np_1 q_1)^2}{np_1 q_1} + \dots$$

$$\frac{\nu_{11}}{n} \rightarrow z_{11} \neq p_1 q_1.$$

Тогда в первом слагаемом будет

$$\frac{n\left(\frac{\nu_{11}}{n} - p_1 q_1\right)^2}{p_1 q_1}.$$

Вывод: если гипотеза не выполняется, то

$$T \xrightarrow{P} +\infty, n \rightarrow \infty.$$

Гипотезы о параметрах нормального распределения

1. Гипотеза о среднем при известной дисперсии.

x_1, \dots, x_n – н.о.р. $N(a, \sigma^2)$

$H = \{a = a_0\}$ – гипотеза

$$T = \sqrt{n} \frac{\bar{x} - a_0}{\sigma} \sim N(0; 1)$$

Схема.

1) Считаем T при выполнении гипотезы H

2) Находим γ_α

Рис.

Упражнение. Выразить γ_α через квантиль для $N(0; 1)$

3) $T \in (-\gamma_\alpha; \gamma_\alpha) \rightarrow$ да

$T \notin (-\gamma_\alpha; \gamma_\alpha) \rightarrow$ нет.

Если $a \neq a_0$

$$T = \sqrt{n} \frac{\bar{x} - a}{\sigma} + \sqrt{n} \frac{a - a_0}{\sigma}$$

$$T \rightarrow +\infty, n \rightarrow \infty$$

2. Гипотеза о среднем при неизвестной дисперсии.

$$T = \frac{(\bar{x} - a_0)\sqrt{n}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{(\bar{x} - a_0)\sqrt{n}/\sigma}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}/\sigma}$$

Утверждение. $\bar{x} - a$ и $\sum_{i=1}^n (x_i - \bar{x})^2$ независимы.

▷ Достаточно доказать, что \bar{x} и $(x_1 - \bar{x}, \dots, x_n - \bar{x})$ независимы. Они совместно гауссовские и достаточно доказать некоррелированность.

$$\text{Cov}(\bar{x}, x_1 - \bar{x}) = \text{Cov}(\bar{x}, x_1) - \text{Cov} \bar{x}, \bar{x} = 0$$

$$\text{Cov}(\bar{x}, x_1) = \frac{1}{n} \text{Cov}(x_1 + \dots + x_n, x_1) = \frac{\sigma^2}{n} = 0$$

$$\text{Cov}(\bar{x}, \bar{x}) = \frac{\sigma^2}{n}. \quad \triangleleft$$

$$T \sim \frac{N(0; 1)}{\sqrt{\frac{1}{n-1} \chi_{n-1}^2}}$$

Определение. Распределением Стьюдента с n степенями свободы называется распределение величины

$$\frac{\xi_0}{\sqrt{\frac{1}{n}(\xi_1^2 + \dots + \xi_n^2)}}$$

где $\xi_0, \xi_1, \dots, \xi_n \sim N(0; 1)$, независимы.

Обозначается

$$t_n \sim \frac{N(0; 1)}{\sqrt{\frac{1}{n} \chi_n^2}}$$

1) Считаем T

2) Находим γ_α :

Рис.

3) $T \in (\gamma_\alpha; \gamma_\alpha) \rightarrow$ принимаем

$T \notin (\gamma_\alpha; \gamma_\alpha) \rightarrow$ отвергаем

У.Госсет (химик) в 1908 г. под псевдонимом “Student” опубликовал работу, в которой определил распределение t_n .

3. Сравнение средних у двух нормальных выборок.

$$x_1, \dots, x_n \sim N(a_x, \sigma^2)$$

$$y_1, \dots, y_m \sim N(a_y, \sigma^2)$$

$$H = \{a_x = a_y\}$$

$$T = \sqrt{\frac{mn}{m+n}} \frac{\bar{x} - \bar{y} / \sigma}{\sqrt{\frac{1}{n+m-2} \left(\underbrace{\sum_{k=1}^n (x_k - \bar{x})^2}_{\chi_{n-1}^2} + \underbrace{\sum_{l=1}^n (y_l - \bar{y})^2}_{\chi_{m-1}^2} \right) / \sigma}} \quad \chi_{n+m-2}^2$$

$$D(\bar{x} - \bar{y}) = \sigma^2 \left(\frac{nm}{n+m} \right)$$

$$T \sim t_{n+m-2}$$

Примечание.

- 1) Критерий Стьюдента в зад.3 работает только если уже известно, что
 - а) наблюдения имеют нормальное распределение
 - б) дисперсии в сериях совпадают.

Замечание. При $n \rightarrow \infty$

$$t_n \Rightarrow N(0; 1)$$

- Упр.** а) Посчитать t_1 – распределение Коши;
 б) t_n .

Различение двух простых гипотез.

Критерий Неймана–Пирсона

Общая постановка: Есть набор наблюдений x_1, \dots, x_n и есть два взаимоисключающих предположения о распределении наблюдений H_0, H_1 . Надо предъявить критерий по $\{x_k\}$, выбирающий гипотезу.

P_0 – распределение выборки при H_0

P_1 – распределение выборки при H_1

Критерий $\longleftrightarrow D \subset \mathbb{R}^n$

$(x_1, \dots, x_n) \in D \rightarrow H_0$

$(x_1, \dots, x_n) \notin D \rightarrow H_1$

Определение. Вероятность $\alpha = P_0(\mathbb{R}^n \setminus D)$ вероятность ошибки 1-го рода.

$\beta = P_1(D)$ – ошибка 2-го рода.

Упражнение. Тест на выявление болезни.

Больных – 1%. Ошибки 1-го и 2-го рода теста равны по 0,1%. Тест получил положительный результат. Какова вероятность того, что человек болен.

Общая ситуация: с уменьшением α β возрастает.

Критерий Неймана–Пирсона. Запишем $T = \frac{p_0}{p_1}$, предполагаем, что распределение T не имеет атомов

(F_T – непрерывна и $\forall \alpha \in [0; 1] \exists C : F_T(C) = \alpha$)

$$D_\alpha^* = \{T \geq C\}$$

Лемма (Неймана–Пирсона). $\forall D \subset \mathbb{R}^n : P_0(\mathbb{R}^n \setminus D) = \alpha$

$$\beta = P_1(D) \geq P_1(D_\alpha^*) =: \beta^*(\alpha)$$

т.е. критерий Неймана–Пирсона является равномерно наиболее мощный среди всех критериев.

Лекция 12

H_0, H_1 – две гипотезы, F_0, F_1 – соответствующие им распределения. $G \subset \mathbb{R}^n$ – критерий.

Ошибка 1-го рода: отвергаем H_0 , когда она верна.

$$P_0(\mathbb{R}^n \setminus G)$$

Ошибка 2-го рода: принимаем H_0 , когда она не верна: $P_1(G)$.

Пусть α – допустимый уровень ошибки 1-го рода. Задача состоит в том, чтобы построить область G , для которой

$$1) P_0(\mathbb{R}^n \setminus G) \leq \alpha$$

$$2) P_1(G) \searrow$$

Считаем, что

F_0 имеет плотность p_0

F_1 имеет плотность p_1 и $p_0, p_1 > 0$.

$\frac{L_1(\vec{x})}{L_0(\vec{x})}$ – отношение функций правдоподобия.

$$G_c = \left\{ \vec{x} : \frac{L_1(x)}{L_0(x)} \geq c \right\}, \quad c > 0$$

$$\begin{aligned} P_1(\vec{x} : \vec{x} \in G_c) &= \int_{G_c} L_1(x) dx = \\ &= \int_{G_c} \frac{L_1(x)}{L_0(x)} L_0(x) dx \geq c P_0\{\vec{x} : \vec{x} \in G_c\}. \end{aligned}$$

Пусть для $\alpha > 0 \exists C_\alpha > 0 : P_0\{\vec{x} : \vec{x} \notin G_c\} = \alpha$.

Теорема (лемма Неймана–Пирсона). *Оптимальный критерий для уровня ошибки 1-го рода α – это область*

$$\overline{G}_{C_\alpha} = \mathbb{R}^n \setminus G_{C_\alpha}$$

▷ $G^* = \overline{G}_{C_\alpha}$. Пусть $G \subset \mathbb{R}^n : P_0(\vec{x} \notin G) = \alpha$. Надо доказать, что

$$P_1(\vec{x} \in G^*) \leq P_1(\vec{x} \in G)$$

Рис.

$$\begin{aligned} P_1(\vec{x} \in G^*) &= \int_{G^*} \frac{L_1}{L_0} L_0 dx \\ &= \int_{G^* \setminus G} \underbrace{\frac{L_1}{L_0}}_{< C_\alpha} L_0 dx \\ &\quad + \int_{G \setminus G^*} \underbrace{\frac{L_1}{L_0}}_{\geq C_\alpha} L_0 dx \\ \int_{G^* \setminus G} \frac{L_1}{L_0} L_0 dx &\leq C_\alpha \int_{G^* \setminus G} L_0 dx \\ \int_{G \setminus G^*} \frac{L_1}{L_0} L_0 dx &\leq C_\alpha \int_{G \setminus G^*} L_0 dx \\ \int_G L_0 dx &= \int_{G^*} L_0 dx = 1 - \alpha \\ &\Downarrow \\ \int_{G^* \setminus G} L_0 dx &= \int_{G \setminus G^*} L_0 dx \end{aligned}$$

и имеет место нужное неравенство. ◁

Пример. Различение гипотез о среднем для нормального распределения.

$$H_0 : N(a_0, 1)$$

$$H_1 : N(a_1, 1), \quad a_1 > a_0.$$

$$L_0(\vec{x}) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - a_0)^2}$$

$$\begin{aligned}
L_1(\vec{x}) &= \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - a_1)^2} \\
\frac{L_1(\vec{x})}{L_0(\vec{x})} &= e^{\frac{1}{2} (\sum_{i=1}^n (x_i - a_0)^2 - \sum_{i=1}^n (x_i - a_1)^2)} = \\
&= e^{\frac{1}{2} (x_1^2 - 2x_1a_0 + a_0^2 + \dots - x_1^2 + 2x_1a_1 - a_1^2 + \dots)} = \\
&= e^{\sum_{i=1}^n x_i(a_1 - a_0) + \frac{n}{2}(a_0^2 - a_1^2)}.
\end{aligned}$$

Область будет вида $\{\sum_{i=1}^n x_i(a_1 - a_0) \geq c\}$

$$\begin{aligned}
P_0 \left\{ \sum_{i=1}^n (x_i - a_0)(a_1 - a_0) \geq c - na_0(a_1 - a_0) \right\} &= \\
= P_0 \left\{ \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n (x_i - a_0)}_{N(0;1)} \geq \frac{c}{\sqrt{n}(a_1 - a_0)} - \sqrt{n}a_0 \right\}.
\end{aligned}$$

Пример (дискретные распределения.)

$1, \dots, N$

$p_1^0, \dots, p_N^0, \quad L_0(x) = \prod_{i=1}^N p_{x_i}^0$

$p_1^1, \dots, p_N^0, \quad \frac{L_1(x)}{L_0(x)} = \prod_{i=1}^n \frac{p_{x_i}^1}{p_{x_i}^0}$

$$P_0 \left\{ \frac{L_1(x)}{L_0(x)} \geq C \right\} \stackrel{?}{=} \alpha$$

α может не достигаться.

Пусть Φ – ступенчатая функция, имеющая скачки в точках c_1, \dots, c_k , в которых она принимает значения $\alpha_1, \dots, \alpha_k$.

Если α такое, что $\alpha_j < \alpha < \alpha_{j+1}$, то выбираем c_{α_j} .

Рандомизация

G_j, G_{j+1}

с вероятностью $\frac{\alpha_{j+1} - \alpha}{\alpha_{j+1} - \alpha_j}$ выбираем G_j ,

с вероятностью $\frac{\alpha - \alpha_j}{\alpha_{j+1} - \alpha_j}$ выбираем G_{j+1} .

Упражнение доделать.

Литература

1. Кендалл, Стьюарт. Статистические выводы и связи.
2. Ивченко, Медведев. Математическая статистика.
3. Худсон. Статистика для физиков.
4. Феллер. Введение в теорию вероятностей.