

June 1, 2016

0.0 Список вопросов к экзамену по математической статистике

1. Случайная выборка и генеральная совокупность.
2. Функции распределения выборки
2. Эмпирические функции распределения
2. Гистограмма
3. Выборочные характеристики. Выборочные моменты
4. Точечные оценки и их свойства
5. Функция правдоподобия. Неравенство Крамера-Рэо
6. Метод максимального правдоподобия: свойства, оценка максимального правдоподобия
7. Метод моментов для точечных оценок
8. Доверительные статистики
9. Интервальные оценки. Доверительные интервалы
10. Интервальные оценки. Доверительные интервалы для дисперсии нормальной генеральной совокупности
11. Асимптотические свойства оценки максимального правдоподобия. Асимптотический доверительный интервал
12. Проверка статистических гипотез. Критерий Нобелья-Парсона проверки простых гипотез
13. Наиболее мощный критерий. Тестерия Нобелья-Парсона
14. Проверка статистических гипотез о параметрах нормального распределения
15. Критерии для сложных гипотез
16. Функция мощности при альтернативе
17. Критерий согласия  $\chi^2$ -Парсона
18. Критерий согласия Колмогорова
19. Критерий однородности Колмогорова-Смирнова
20. Критерий однородности  $\chi^2$

0.1 Случайная выборка, генеральная совокупность, функция распределения выборки

**Def. 1.** Выборка Пусть эксперимент состоит в проведении  $n$  испытаний, результат  $j$ -го из которых является случайной величиной  $X_j$  ( $\Omega_j \rightarrow X_j$ ). Кортеж из этих случайных величин (случайный вектор)  $X = (X_1, \dots, X_n)$  называется (случайной) выборкой, а г.в.  $X_j$  называется элементом выборки. А значение  $x = (x_1, \dots, x_n) \in X(\omega)$  называется реализацией выборки. Далее всегда, если не указано иное, случайные величины будут обозначаться заглавными буквами, а их реализации соответствующими строчными. Далее  $X_j$  понимается как независимый.

**Def. 2.** Выборочное пространство Выборочным пространством называется измеримое пространство  $(X, A)$ , где  $X = \{X(\omega); \omega \in \Omega\}$  есть множество возможных значений выборки, а  $\mathcal{F} = \sigma$ -алгебра в  $X$ .

Особенно важен случай, когда случайные величины  $X_j$  являются независимыми и имеют распределение одной случайной величины  $\xi$ . Этот случай соответствует повторению  $n$  раз одного эксперимента, описанного случайной величиной  $\xi$ .

**Def. 3.** Генеральная совокупность Генеральной совокупностью называют распределение  $\mathcal{L}(\xi)$  случайной величины  $\xi$ . Оно может быть задано, например, множеством возможных значений г.в.  $\xi$  и её функцией распределения. При этом  $X$  называют выборкой из (генеральной совокупности)  $\mathcal{L}(\xi)$ .

**Def. 4.** Функции распределения выборки  $X \in \mathcal{L}(\xi)$

$$F_X(x) = P(X \leq x) = \prod_{j=1}^n P(X_j \leq x_j) = \prod_{j=1}^n F_{X_j}(x_j)$$

0.2 Эмпирическая функция распределения, гистограмма

Пусть  $A \subset \mathbb{R}$  событие, произошедшее в ходе испытания с вероятностью  $P(A) = p$ , и пусть эксперимент состоит в проведении  $n$  таких независимых испытаний. Тогда

$$\Omega = \prod_{j=1}^n \Omega_{A_j}$$

А случайная величина

$$X_j = I_{\{\omega | \omega \in A\}} = \begin{cases} 1, & \omega_j \in A \\ 0, & \omega_j \notin A \end{cases}$$

является индикатором того, что в ходе  $j$ -го испытания случилось событие  $A$ .

Пусть г.в.  $k = \sum_{j=1}^n X_j$  — число успехов.  $X_j$  — число успехов.  $A$  в ходе эксперимента.

Выводы г.в.  $p_k^* = \frac{1}{n} \sum_{j=1}^n X_j$ .

Оценку  $E p_k^* = p$ .

Кроме того, из ЗБЧ в форме Бернулли следует

$$\lim_{n \rightarrow \infty} P[|p_k^* - p| < \varepsilon] = 1 \quad \forall \varepsilon > 0$$

Таким образом, значение случайной величины  $p_k^*$  можно считать приближением истинной величины  $p$ .

Пусть теперь  $X = (X_1, \dots, X_n)$  — выборка, образуя  $n$  из генеральной совокупности  $\mathcal{L}(\xi)$ ,  $x = (x_1, \dots, x_n)$  — реализация.

**Def. 5.** Порядковые статистики Каждой реализации  $x$  можно сопоставить в соответствие сго перестановку  $x_{(1)} \leq \dots \leq x_{(n)}$ .  $j$ -й порядковой статистикой называют случайная величина  $X_{(j)}$ , при каждой реализации  $X(\omega) = x$ , принимает значение  $x_{(j)}$  ( $x_{(j)}$ ).

**Def. 6.** Вариационный рас Случайный вектор  $(X_{(1)}, \dots, X_{(n)})$  называется вариационным расом.

**Def. 7.** Эмпирическая функция распределения Для каждого  $t \in \mathbb{R}$  заданную случайную величину  $\mu_n(t)$ , равную количеству элементов выборки  $X$ , значения которых не превосходят  $t$ :

$$\mu_n(t) = \sum_{j=1}^n I_{\{x_j \leq t\}}$$

Эмпирической функцией распределения, построенной по выборке  $X$ , называют случайную функцию  $F_n : t \mapsto \mathcal{L}^0(\Omega)$

$$F_n(x) = \frac{1}{n} \mu_n(t)$$

Её значение в точке  $t$  является случайной величиной, сопоставленной по вероятности  $P(t)$  теоретической функции распределения EDF можно переписать с помощью функции Хемпбелла (Hemphill):

$$H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$$

$$F_n(t) = \frac{1}{n} \sum_{j=1}^n H(t - X_{(j)})$$

**Def. 8.** Гистограмма Ресольбы область значений г.в.  $\xi$  на равные интервалы  $\Delta_n$ , и для каждого  $\Delta_n$  подставляем число  $n_k$  элементов  $x_j$  вектора  $x$ , попавших в  $\Delta_n$ ,  $n = \sum_{k=1}^m n_k$ .

Построим график ступенчатой функции

$$t \mapsto \frac{n_k}{n \Delta_n}, \quad t \in \Delta_n, b_i = [\Delta_n]$$

Полученный график (при желании, само отображение) называется Гистограммой, построенной по данной реализации выборки. Соединив середины смежных отрезков этого графика. Полученная ломаная называется пилочным частот.

С увеличением  $\max\{b_i\}$ , гистограмма в пилочном частот всё более точно приближает вероятности попадания в каждый из интервалов разбиения.

Contents

1

0.3 Выборочные характеристики. Выборочные моменты

Пусть  $X = (X_1, \dots, X_n)$  — выборка из  $\mathcal{L}(\xi)$ ,  $F$  и  $F_n$  — соответственно теоретическая и эмпирическая функции распределения.

Вектор характеристике  $j$  случайной величины  $\xi$

$$\theta = \int_{\mathbb{R}} g(t) dF(t)$$

можно поставить в соответствие статистический аналог — случайную величину  $G$ :

$$G = \int_{\mathbb{R}} g(t) dF_n(t) = \frac{1}{n} \sum_{j=1}^n g \circ X_j$$

$$G(\omega) = \int_{\mathbb{R}} g(x) d((F_n(t))(\omega)) = \frac{1}{n} \sum_{j=1}^n g(x_j)$$

Выборочные моменты  $k$ -го порядка называются статистический аналог характеристики  $\alpha_k = \mathbb{E} x^k = \int_{\mathbb{R}} t^k dF(t)$ :

$$A_k = \frac{1}{n} \sum_{j=1}^n X_j^k$$

$\bar{X} = A_1$  называют выборочным средним.

Выборочные центральным моментом  $k$ -го порядка называют случайную величину  $M_k$  — статистический аналог характеристики  $\mu_k = \mathbb{E}(\xi - \mathbb{E} \xi)^k = \int_{\mathbb{R}} (t - \alpha_1)^k dF(t)$

$$M_k = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^k$$

$M_2$  называют выборочной дисперсией

NB 1. Выборочное среднее является несмещённой оценкой математического ожидания

$$\mathbb{E} X = \frac{1}{n} \sum_{j=1}^n \mathbb{E} X_j = \frac{n \alpha_1}{n} = \alpha_1$$

0.4 Точечные оценки

Пусть, который процесс описывается вероятностной моделью  $(\Omega, A, P)$ , где  $\Omega$  — пространство элементарных событий,  $A \subset 2^{\Omega}$  —  $\sigma$ -алгебра событий,  $P : A \rightarrow [0, 1]$  — вероятностная мера, а проводимый эксперимент соответствует случайному значению  $\xi \in \mathcal{D}^1$ ,  $\theta$  функцией распределения  $F$ .

Рассмотрим задачу определения распределения случайной величины, в случае когда известно, что её функция распределения  $F$  принадлежит некоторому классу распределений, зависящих от параметра

$$F \in \mathcal{F} = \{F_\theta; \theta \in \Theta\}$$

где  $\Theta$  — множество значений некоторого параметра  $\theta$ . То есть известно, что распределение определяется некоторым конкретным значением  $\theta$ , и задача сводится к его оценке.

Пусть  $X = (X_1, \dots, X_n)$  — выборка из  $\mathcal{L}(\xi)$ . Говорят, что пара  $(X, \mathcal{F})$  задёт "статистическую модель".

**Def. 9.** Статистика Статистикой называются случайная величина — композиция  $g \circ X$  некоторой (вообще говоря борелевской) функции  $g$  и выборки  $X$

$$(g \circ X)(\omega) = g(x)$$

**Def. 10.** Точечная оценка параметра  $\theta$  есть статистика  $T \circ X$  (или для простоты сама функция  $T$ ), реализацию  $(T \circ X)(\omega) = T(x)$  которой принимают за приближённое значение параметра  $\theta$ .

0.4.1 Характеристики оценок

**Def. 11.** Несмещённость (unbiasedness) Несмещённой называют такую оценку  $T$ , что её математический ожидаемый является истинный параметр  $\theta$

$$\mathbb{E}(T \circ X) = \theta$$

**Def. 12.** Состоятельность Оценка  $T$  называется состоятельной, если она сходится по вероятности к оцениваемому параметру:

$$\lim_{n \rightarrow \infty} P(\omega \in \Omega : |(T \circ X)(\omega) - \theta| < \varepsilon) = \lim_{n \rightarrow \infty} P(|T(x) - \theta| < \varepsilon) = 1$$

**Def. 13.** Оптимальность Оценка  $T_0$  называется оптимальной в классе несмещённых оценок  $\mathcal{T}$ , если среди всех оценок класса  $\mathcal{T}$ , оценка  $T_0$  имеет минимальную дисперсию, то есть для любого  $T \in \mathcal{T}$

$$D(T_0 \circ X) \leq D(T \circ X)$$

Оценка называется оптимальной, если она оптимальна в классе всех несмещённых оценок