

June 2, 2016

0.0 Список вопросов к экзамену по математической статистике

1. Случайная выборка и генеральная совокупность.
2. Функции распределения выборки
1. 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 оценок максимального правдоподобия
7. Метод моментов для точечных оценок
8. Достоверные статистики
9. Интервальные оценки. Доверительные интервалы
10. Интервальные оценки. Доверительные интервалы для дисперсии нормальной генеральной совокупности
11. Асимптотические свойства оценок максимального правдоподобия. Асимптотический доверительный интервал
12. Проверка статистических гипотез. Критерий Нобелья-Пирсона проверки простых гипотез
13. Наиболее мощный критерий. Тестерия Нобелья-Пирсона
14. Проверка статистических гипотез о параметрах нормального распределения
15. Критерии для сложных гипотез
16. Функция мощности при альтернативе
17. Критерий согласия  $\chi^2$ -Пирсона
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.** Случайная выборка (sample) Пусть эксперимент состоит в проведении  $n$  испытаний, результат  $j$ -го из которых является случайной величиной  $X_j$ ,  $\Omega_j \ni \omega \rightarrow X_j$ .  
Кортеж из этих случайных величин (случайный вектор)  $X = (X_1, \dots, X_n)$  называется (случайной) выборкой, а г.с.в.  $X_j$  называется элементами выборки.  
А значение  $x = (x_1, \dots, x_n) = X(\omega)$  называется реализацией выборки.  
Далее всегда, если не указано иное, случайные величины будут обозначаться заглавными буквами, а их реализации соответствующими строчными.  
Далее  $X_j$  предполагается независимыми.  
**Def. 2.** Выборочное пространство (sample space) Выборочным пространством называется измеримое пространство  $(X, \mathcal{A})$ , 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 (population) Генеральной совокупностью называют распределение  $L(\xi)$  случайной величины  $\xi$ . Оно может быть задано, например, множеством возможных значений г.с. в  $\Omega$  и её функцией распределения.  
При этом  $X$  называют выборкой из (генеральной совокупности)  $L(\xi)$ .

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

$$F_X(x) = P(X \leq x) = \prod P(X_j \leq x_j) = \prod F_{X_j}(x_j)$$

0.2 Эмпирическая функция распределения, гистограмма

Пусть  $A \subset \mathbb{R}$  события, происходящие в ходе испытания с вероятностями  $P(A) = p$ , и пусть эксперимент состоит в проведении  $n$  таких независимых испытаний.  
Тогда 
$$\Omega = \prod_{j=1}^n \Omega_A$$
$$X_j = I_A(\omega) \in A = \begin{cases} 1, & \omega_j \in A \\ 0, & \omega_j \notin A \end{cases}$$
А случайная величина  
$$X_j = I_A(\omega) \in A = \begin{cases} 1, & \omega_j \in A \\ 0, & \omega_j \notin A \end{cases}$$
является индикатором того, что в ходе  $j$ -го испытания случилось событие  $A$ .  
Пусть г.с.в.  $k = \sum_{j=1}^n X_j$  — число проявлений  $A$  в ходе эксперимента.  
Выводы г.с.в.  $p_k^* = \frac{1}{n} \sum_{j=1}^n X_j$ .  
Очевидно  $E p_k^* = p$ .  
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 приближенной оценкой величины  $p$ .  
Пусть теперь  $X = (X_1, \dots, X_n)$  — выборка, образуя  $n$  из генеральной совокупности  $L(\xi)$ ,  $x = (x_1, \dots, x_n)$  — реализация.  
**Def. 5.** Порядокности. Каждой реализации  $x$  можно сопоставить в соответствие его 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  $X_{(j)}(\omega) = x_{(j)}$ .  
**Def. 6.** Вариационный ряд. Случайный вектор  $(X_{(1)}, \dots, X_{(n)})$  называется вариационным рядом.  
**Def. 7.** Эмпирическая функция распределения. Для каждого  $t \in \mathbb{R}$  задана случайную величину  $\mu_n(t)$ , равную количеству элементов выборки  $X$ , значения которых не превосходят  $t$ .

$$\mu_n(t) = \sum I_{(-\infty, t]}(X_j)$$

Эмпирической функцией распределения, построенной по выборке  $X$ , называют случайную функцию  $F_n : t \mapsto L_n^*(\Omega)$

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

Её значение в точке  $t$  является случайной величиной, создаваемой по вероятности  $p$  значением  $F(t)$  теоретической функции распределения EDF можно переписать с помощью функции Хемминга (Hamming):

$$H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$$
$$F_n(t) = \frac{1}{n} \sum_{j=1}^n H(t - X_{(j)})$$

**Def. 8.** Гистограмма. Разобьем область значений г.с. в на равные интервалы  $\Delta_i$ , и для каждого  $\Delta_i$  подсчитаем число  $n_i$  элементов  $x_j$  вектора  $x$ , попавших в  $\Delta_i$ ,  $n = \sum n_i$ .  
Построим график ступенчатой функции 
$$t \mapsto \frac{n_i}{n \Delta_i}, \quad t \in \Delta_i, b_i = [\Delta_i]$$

Полученный график (при желании, само отображение) называется Гистограммой, построенной по данной реализации выборки.  
Соедини середины смежных отрезков этой графика. Полученная ломаная называется палингом частот.  
С умножением  $\max\{b_i\}$ , гистограмма в палингом 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  $L(\xi)$ ,  $F$  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  $G$ :

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

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

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.  
Выборочным 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

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

2

0.4 Точечные оценки

Пусть  $\omega$  который процесс описывается вероятностной моделью  $(\Omega, \mathcal{A}, P)$ , где  $\Omega$  — пространство элементарных событий,  $\mathcal{A} \subset 2^{\Omega}$  —  $\sigma$ -алгебра событий,  $P : \mathcal{A} \rightarrow [0, 1]$  — вероятностная мера, а произвольный эксперимент соответствует случайной величине  $\xi \in L^2$ , с 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 известным значением  $\theta$ , и задача сводится к его оценке.  
Пусть  $X = (X_1, \dots, X_n)$  — выборка из  $L(\xi)$ . 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  $T = \tau \circ X : \Omega \rightarrow \Theta$ , реализацию  $T(\omega) = \tau(x)$  который принимают за приближённое значение параметра  $\theta$

0.4.1 Характеристики оценок

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

$$\mathbb{E} T = \theta$$

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

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

**Def. 13.** Оптимальность (efficiency) Оценка  $T_0$  называется оптимальной в классе несмещённых оценок  $T$ , если среди всех оценок класса  $T$  оценка  $T_0$  имеет наименьшую дисперсию, то есть для любого  $T \in T$

$$DT_0 \leq DT$$

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

**Тем. 1.** Единственность оптимальной оценки Если две несмещённых оценки  $T_1, T_2$  параметра  $\theta$  оптимальны, то они равны почти-всюду  $T_1 \stackrel{a.s.}{=} T_2$ :

$$P(T_1 \neq T_2) = 0$$

3

4