

June 2, 2016

0.0 Список вопросов к экзамену по математической статистике

1. Случайная выборка и генеральная совокупность.
2. Функции распределения выборки
2. Эмпирическая функция 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. Достаточные 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 Нобельма-Парсона проверки простых гипотез
13. Наиболее мощный критерий. Тестерма Нобельма-Парсона
14. Проверка статистических гипотез о параметрах нормального распределения
15. Критерии для сложных гипотез
16. Функция мощности при альтернативе
17. Критерий согласия χ^2 -Парсона
18. Критерий согласия Колмогорова
19. Критерий однородности Колмогорова-Смирнова
20. Критерий однородности χ^2

0.1 Случайная выборка, генеральная совокупность, функция распределения выборки

Def. 1. Выборочная выборка (sample) Пусть эксперимент состоит в проведении n испытаний, результат j -го из которых является случайной величиной X_j . $\Omega_1 \rightarrow X_1$.
Кортеж из этих случайных величин (случайный вектор) $X = (X_1, \dots, X_n)$ называется (случайной) выборкой, а г.т.в. X_j называется элементом 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 рассматриваются 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} — σ -алгебра в X .
Особенно назовем случай, когда случайные величины X_j являются независимыми и имеют распределение одной случайной величины ξ . Этот случай соответствует повторению n раз одного эксперимента, описанного случайной величиной ξ .
Def. 3. Генеральная совокупность (population) Генеральной совокупностью называют распределение $L(\xi)$ случайной величины ξ . Оно может быть задано, например, множеством возможных значений г.т. ξ и об функции распределения.
При этом X называют выборкой из (генеральной совокупности) $L(\xi)$.
Def. 4. Функция 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

Def. 1. $A \subset \mathbb{R}$ событие, произошедшее в ходе испытания с вероятностью $P(A) = p$, и пусть эксперимент состоит в проведении n таких независимых испытаний.
Тогда
$$\Omega = \prod_{j=1}^n \Omega_k$$
$$X_j = I_{\{\omega : \omega_j \in A\}} = \begin{cases} 1, & \omega_j \in A \\ 0, & \omega_j \notin A \end{cases}$$
$$X_j = I_{\{\omega : \omega_j \in A\}} = \begin{cases} 1, & \omega_j \in A \\ 0, & \omega_j \notin A \end{cases}$$
 является индикатором того, что в ходе j -го испытания случилось событие A .
Пусты г.т.г. $k = \sum_{j=1}^n X_j$ — число проявлений A в ходе эксперимента.
Выводы г.т.г. $p_k = \frac{1}{2^n} \sum_{j=1}^n X_j$.
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, добытая из генеральной совокупности $L(\xi)$, $x = (x_1, \dots, x_n) =$ реализация.
Def. 5. Порядковые статистики Каждой реализации x можно сопоставить в соответствие его перестановку $x_{(1)} \leq \dots \leq x_{(n)}$.
 j -й порядковой статистикой назовем 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 .

$$\mu_n(t) = \sum I_{\{X_j \leq t\}}$$

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

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

Ее значение в точке t является случайной величиной, сопоставленной по вероятности к значению $F(t)$ 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.т. ξ на равные интервалы Δ_n , и для каждого Δ_n подберем число n_k элементов x_j вектора x , попавших в Δ_n , $n = \sum n_k$.
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.
Соединим середины смежных ступенек этого графика. Полученная ломаная называется палингом частот.
С увеличением $\max\{b_i\}$, гистограмма в палингом частот все более точно приближает вероятности попадания в каждый из интервалов разбиения.

Contents

1

2

3

4

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.
Возьмем характеристику j случайной величины ξ

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

можно поставить в соответствие статистический аналог — случайную величину G :

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

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

Выборочный моментом k -го порядка называется 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.
Выборочный центральный моментом k -го порядка назовем 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. Выборочные средние являются несмещенной оценкой математического ожидания

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

0.4 Точечные оценки

Пусть эксперимент процесс описывается вероятностной моделью (Ω, \mathcal{A}, P) , где Ω — пространство элементарных событий, $\mathcal{A} \subset 2^{\Omega}$ — σ -алгебра событий, $P : \mathcal{A} \rightarrow [0, 1]$ — вероятностная мера, а проводимый эксперимент соответствует случайной величине $\xi \in L^2$, с функцией распределения F .

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

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

где Θ — множество значений параметра θ . То есть известно, что распределение определяется некоторым известным известным известным значением θ_k параметра θ , и задача сводится к его оценке.

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

Def. 9. Статистика Статистикой называется 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 θ есть статистика $T = \tau \circ X : \Omega \rightarrow \Theta$, реализацию $T(\omega) = \tau(x)$ который принимают за приближенное значение параметра θ

0.4.1 Характеристики оценок

Def. 11. Несмещенность (unbiasedness) Несмещенной называют такую оценку T , что её математическое ожидание является искомым параметром θ

$$\mathbb{E} T = \int_{\mathbb{R}} \tau(x) f(x; \theta_0) dx = \theta_0$$

Def. 12. Состоятельность (consistency) Оценка T_n называется состоятельной, если она сходится по вероятности к искомому параметру:

$$\lim_{n \rightarrow \infty} P[\omega \in \Omega : |T(\omega) - \theta_0| < \varepsilon] = \lim_{n \rightarrow \infty} P[|\tau(x) - \theta_0| < \varepsilon] = 1$$

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

$$D T_0 \leq D T$$

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

Тем. 1. Единственность оптимальной оценки Если две несмещенные оценки T_1, T_2 параметра θ оптимальны, то они равны почти-всюду $T_1 = T_2$:

$$P[T_1 \neq T_2] = 0$$