Конспекты к экзамену по математической статистике

June 3, 2016

Contents

0.0 Список вопросов к экзамену по математической статистике

1. 1. Случайная выборка и генеральная совокупность
   2. Функция распределения выборки
2. 1. Эмпирическая функция распределения
   2. Гистограмма
3. Выборочные характеристики. Выборочные моменты
4. Точечные оценки и их свойства
5. Функция правдоподобия. Неравенство Крамера-Рао
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 интервал для дисперсии нормальной генеральной совокупности
11. Асимптотические свойства оценки максимального правдоподобия.

    Асимптотический доверительный интервал
12. Проверка статистических гипотез.

    Критерий Неймана-Пирсона проверки простых гипотез
13. Наиболее мощный критерий. Теорема Неймана-Пирсона
14. Проверка статистических гипотез о параметрах нормального распределения
15. Критерии для сложных гипотез
16. Функция мощности при альтернативе
17. Критерий согласия χ^2-Пирсона
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. Выборка (sample) Пусть эксперемент состоит в проведении n испытаний, результат j-го из которых является случайной величиной $X_j : \Omega_j \to \mathcal{X}_j$.

Кортёж из этих случайных величин (случайный вектор) $X = (X_1, \ldots, X_n)$ называется (случайной) выборкой, а r.v. X_j называются элементами выборки

А значение $x = (x_1, \ldots, x_n) = X(\omega)$ называется реализацией выборки

Далее всегда, если не указано иное, случайные величины будут обозначаться заглавными буквами, а их реализации соответствующими строчными

Далее X_j полагаются независимыми

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

Особенно важен случай, когда случайные величины X_j являются независимыми и имеют распределение одной случайной величины ξ. Этот случай соответствует повторению n раз одного эксперемента, описываемого случайной величиной ξ

Def. 3. Генеральная совокупность (population) Генеральной совокупностью называют распределение $\mathcal{L}(\xi)$ случайной величины ξ

Оно может быть задано, например, множеством возможных значений r.v. ξ и её функцией распределения

При этом X называют выборкой из (генеральной совокупности) $\mathcal{L}(\xi)$

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

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

0.2 Эмпирическая функция распределения, гистограмма

Пусть $A \subset \Omega_0$ событие, происходящее в ходе испытания с вероятностью $\mathbb{P}A = p$, и пусть эксперимент состоит в проведении n таких независимых испытаний

Тогда

$$\Omega = \prod_{j=1}^{n} \Omega_0$$

А случайная величина

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

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

Пусть r.v. $k = \sum_{j=1}^{n} X_j$ — число проявлений A в ходе эксперимента.
Введём r.v. $p_n^* = \frac{1}{n} \sum_{j=1}^{n} X_j$.
Очевидно $\mathbb{E}p_n^* = p$.
Кроме того, из ЗБЧ в форме Бернулли следует

$$\lim_{n\to\infty} \mathbb{P}\{|p_n^* - p| < \varepsilon\} = 1 \quad \forall\varepsilon > 0$$

Таким образом, значение случайной величины p_n^* можно считать приближённой оценкой величины p

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

Def. 5. Порядковые статистики Каждой реализации x можно сопоставить в соответствие его перестановку $x_{(1)} \leq \cdots \leq x_{(n)}$,

j-й порядковой статистикой назвается случайная величина $X_{(j)}$, при каждой реализации $X(\omega) = x$, принимает значение $X_{(j)}(\omega) = x_{(j)}$

Def. 6. Вариационный ряд Случайный вектор $(X_{(1)}, \ldots, X_{(n)})$ называется вариационным рядом

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

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

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, сходящейся по вероятности к значению $F(t)$ теоретической функции распределения

EDF можно перезаписать с помощью функции Хевисайда (Heaviside):

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

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

Def. 8. Гистограмма Разобьём область значений r.v. ξ на равные интервалы Δ_i, и для каждого Δ_i подсчитаем число n_i элементов x_j вектора x, попавших в Δ_i, $n = \sum n_i$.

Построим график ступенчатой функции

$$t \mapsto \frac{n_i}{nh_i}, \quad t \in \Delta_i, h_i = |\Delta_i|$$

Полученный график (при желании, само отображение) называется Гистограммой, построенной по данной реализации выборки

Соединим середины смежных отрезков этого графика. Полученная ломанная называется полигоном частот

С уменьшением $\max\{h_i\}$, гистограмма и полигон частот всё более точно приближают вероятности попадания в каждый из интервалов разбиения

0.3 Выборочные характеристики. Выборочные моменты

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

Всякой характиристике $\tilde{g}$ случайной величины ξ

$$\tilde{g} = \int_{\mathbb{R}} g(t) \mathrm{d}F(t)$$

можно поставить в соответствие статистический аналог — случайную величину G:

$$G = \int_{\mathcal{X}} g(x) \mathrm{d}F_n(x) = \frac{1}{n} \sum_{j=1}^{n} g \circ X_j$$

$$G(\omega) = \int_{\mathcal{X}} g(x) \mathrm{d}\left((F_n(x))(\omega)\right) = \frac{1}{n} \sum_{j=1}^{n} g(x_j)$$

Выборочным моментом k-го порядка называется статистический аналог характеристики $\alpha_k = \mathbb{E}\xi^k = \int_{\mathbb{R}} t^k \mathrm{d}F(t)$:

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

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

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

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

M_2 называют выборочной дисперсией

NB 1. Выборочное среднее является несмещённой оценкой математического ожидания

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

0.4 Точечные оценки

Пусть некоторый процесс описывается вероятностной моделью $(\Omega, \mathcal{A}, \mathbb{P})$, где Ω — пространство элементарных событий, $\mathcal{A} \subset 2^{\Omega}$ — σ-алгебра событий, $\mathbb{P} : \mathcal{A} \to [0,1]$ — вероятностная мера, а проводимый эксперимент соответствует случайной величине $\xi \in \mathcal{L}^0$, 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 Θ — множество значений параметра θ. То есть известно, что распределение определяется некоторым неизвестным истинным значением θ_0 параметра θ, и задача сводится к его оценке.

Пусть $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi)$. Говорят, что пара $(\mathcal{X}, \mathcal{F})$ задаёт “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 $T = \tau \circ X : \Omega \to \Theta$, реализацию $T(\omega) = \tau(x)$ которой принимают за приближённое значение параметра θ

0.4.1 Характеристики оценок

Def. 11. Несмещённость (unbiasedness) Несмещённой называют такую оценку T, что её математическим ожиданием является искомый параметр θ:

$$\mathbb{E}T = \int_{\mathcal{X}} \tau(x) f(x; \theta_0) \mathrm{d}x = \theta_0$$

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

$$\lim_{n\to\infty} \mathbb{P}\{\omega \in \Omega; |T(\omega) - \theta_0| < \varepsilon\} = \lim_{n\to\infty} \mathbb{P}\{|\tau(x) - \theta_0| < \varepsilon\} = 1$$

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

$$\mathbb{D}T_0 \le \mathbb{D}T$$

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

Thm. 1. *Единственность оптимальной оценки Если две несмещённые оценки* T_1, T_2 *параметра* θ *оптимальны, то они* равны почти-всюду $T_1 \overset{\mathbb{P}}{=} T_2$*:*

$$\mathbb{P}\{T_1 \neq T_2\} = 0$$

0.5 Функция правдоподобия. Неравенство Крамера-Рао

Зададим класс допустимых распределений r.v. $\xi : \Omega_0 \to \mathcal{X}_0 \subset \mathbb{R}$ функцией двух переменных

$$f : \mathbb{R} \times \Theta \to [0,1]$$

при каждом фиксированном θ являющейся плотностью распределения вероятностей $f_\theta : \mathbb{R} \to [0,1]$ соответствующего значению θ параметра

Пусть $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi)$, $x = X(\omega)$ — реализация выборки. Функцией $f_s : \mathcal{X} \times \Theta \to [0,1]$ будем обозначать плотность распределения вероятностей выборки $f_s(x;\theta) = \prod_{j=1}^n f(x_j;\theta)$

Def. 14. Функция правдоподобия (likelihood fuction) При фиксированном $x \in \mathcal{X}$ функция $L : \theta \mapsto f_s(x;\theta)$ называется функцией правдоподобия. При данной реализации x, значение $L(\theta)$ характеризует вероятность получения такой реализации при таком значении параметра, i.e. "правдоподобие" этого значения параметра.

Далее будем считать, что при любом x отображение f_s дифференцируемо по θ ($\iff$ f дифференцируемо по θ)

Def. 15. Вклад (score) выборки Пусть

$$u = \frac{\partial \ln f_s}{\partial \theta} : \mathcal{X} \times \Theta \to \mathbb{R}$$

— частная производная функции логарифма правдоподобия. Вкладом выборки X называется случайная функция[1] $U : \Theta \to \mathcal{L}^0$

$$U(\theta)(\omega) = u(X(\omega);\theta) = u(x;\theta)$$

- Вклад характеризует чувствительность плотности распределения выборки к изменению значения параметра θ
- Кроме того, реализация вклада (выборки в данное значение параметра) характеризует "степень несоответствия" данного значения параметра полученной реализации выборки.

Def. 16. Регулярная статистическая модель Статистическая модель, позволяющая дифференцировать (всякие $\int L$ и вообще всё что вздумается) по θ, переставлять операторы интегрирования и дифференцирования, и разрешающая прочий матан называется регулярной

Далее рассматриваются регулярные модели

[1]Далее случайные функции (аналогично последовательности, векторы) могут рассматриваться и как $\mathbb{R} \to \mathcal{L}^0$ отображения в пространство случайных величин, и как $\Omega \to \mathcal{L}^0(\mathbb{R},\mathbb{R})$ отображение из пространства элементарных событий в пространство измеримых функций, и как $\Omega \times \mathbb{R} \to \mathbb{R}$ в зависимости от контекста и потребностей

Thm. 2. *Свойства функции правдоподобия и вклада*

$$\mathbb{E}_\theta L(\theta) = \int_{\mathcal{X}} f_s(x;\theta)\mathrm{d}x = 1 \quad \forall\theta\in\Theta$$

$$\mathbb{E}_\theta U(\theta) = 0 \quad \forall\theta\in\Theta$$

То есть ожидаемый вклад выборки в истинное значение параметра — 0.
Это соответствует ожиданию достижения максимума функции (логарифма) правдоподобия в точке θ, и минимальному "несоответствию" между предполагаемым значением параметра и полученной реализацией выборки.

Proof. Первое равенство естественно.

$$\mathbb{E}_\theta U(\theta) = \int_{\mathcal{X}} \frac{\partial \ln f_s}{\partial\theta}(x;\theta) f_s(x;\theta)\mathrm{d}x = \int_{\mathcal{X}} \frac{\partial f_s}{\partial\theta}(x;\theta)\frac{1}{f_s(x;\theta)} f_s(x;\theta)\mathrm{d}x =$$
$$= \int_{\mathcal{X}} \frac{\partial f_s}{\partial\theta}(x;\theta)\mathrm{d}x = \frac{\partial}{\partial\theta}\int_{\mathcal{X}} f_s(x;\theta)\mathrm{d}x = \frac{\partial}{\partial\theta}1 = 0$$

□

Def. 17. Информация Фишера *Информацией Фишера* о параметре θ содержащейся в выборке X называется функция $\mathcal{I}:\Theta\to\mathbb{R}_+$ (или её значение):

$$\mathcal{I}(\theta) = \mathbb{D}_\theta U(\theta) = \mathbb{E}_\theta U^2(\theta) = \int_{\mathcal{X}} u(x;\theta) f(x;\theta)\mathrm{d}x$$

А функции $\mathcal{I}_j$

$$\mathcal{I}_j(\theta) = \mathbb{D}_\theta \frac{\partial \ln f(X_j;\theta)}{\partial\theta} = \int_{\mathbb{R}} \left(\frac{\partial \ln f(t;\theta)}{\partial\theta}\right)^2 f(t;\theta)\mathrm{d}t$$

количеством информации о параметре θ, содержащейся в j-м наблюдении

$$\mathcal{I}(\theta) = \sum_{j=1}^n \mathcal{I}_j(\theta) = n\mathcal{I}_1(\theta)$$

Информация Фишера представляет собой ожидание (в предположении о истинности значения θ) некоего абсолютного значения несоответствия значения θ параметра полученной реализации выборки

Thm. 3. *Неравенство Крамера-Рао Для любой несмещённой оценки T параметра θ справедливо*

$$\mathbb{D}_\theta T \geq \frac{1}{\mathcal{I}(\theta)}$$

Proof. $T = \tau\circ X$ — несмещённая оценка. Значит

$$\mathbb{E}_\theta T = \int_{\mathcal{X}} \tau(x) f_s(x;\theta)\mathrm{d}x = \theta \qquad \Big| \frac{\partial}{\partial\theta}$$

$$
\begin{aligned}
1 &= \frac{\partial}{\partial\theta}\int_{\mathcal{X}}\tau(x)f_s(x;\theta)\mathrm{d}x = \int_{\mathcal{X}}\tau(x)\frac{\partial}{\partial\theta}f_s(x;\theta)\mathrm{d}x = \int_{\mathcal{X}}\tau(x)\frac{\partial\ln f_s}{\partial\theta}(x;\theta)f_s(x;\theta)\mathrm{d}x \\
&= \mathbb{E}_\theta(TU(\theta)) = \mathbb{E}_\theta((T-\theta)(U(\theta)-0)) + \underbrace{\theta\mathbb{E}_\theta U(\theta)}_{=0} = \mathrm{cov}(T,U(\theta)) \le \sqrt{\mathbb{D}_\theta T\mathbb{D}_\theta U(\theta)}
\end{aligned}
$$

$$
\mathbb{D}_\theta T \ge \frac{1}{\mathbb{D}_\theta U(\theta)} = \frac{1}{\mathcal{I}(\theta)}
$$

□

Если существует оценка, для которой достигается эта нижняя граница, то, очевидно, эта оценка является оптимальной (и единственной)

Thm. 4. *Если нижняя граница Крамера-Рао достигается, то* $1 = \mathrm{cov}(T,U(\theta)) = \mathbb{D}_\theta T\mathbb{D}_\theta U(\theta)$, *i.e.*

$$
\mathrm{corr}(T,U(\theta)) = \frac{\mathrm{cov}(T,U(\theta))}{\mathbb{D}_\theta T\mathbb{D}_\theta U(\theta)} = 1
$$

Что возможно тогда и только тогда, когда T *и* U *линейно-зависимы, i.e.* $\forall\theta\in\Theta$ *существуют константы* $a(\theta), b(\theta)$, *такие что*

$$
\mathbb{P}\{T \ne a(\theta)U(\theta) + b(\theta)\} = 0
$$

Причём $\forall\theta\in\Theta$

$$
\mathbb{D}_\theta T = \mathbb{D}_\theta(a(\theta)U(\theta)+b(\theta)) = \mathbb{D}_\theta(a(\theta)U(\theta)) = a^2(\theta)\mathbb{D}_\theta U(\theta) = \frac{1}{\mathcal{I}(\theta)}
$$

$$
a(\theta) = \pm\frac{1}{\mathbb{D}_\theta U(\theta)}
$$

$$
\theta = \mathbb{E}_\theta T = b(\theta)
$$

$$
T = \frac{1}{\mathbb{D}_\theta U(\theta)}U(\theta) + \theta
$$

0.6 Оценки максимального правдоподобия

Пусть $X = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(\xi) = \{F_\theta; \theta \in \Theta\}$. При каждом $x \in \mathcal{X}$, $L : \theta \to [0, 1]$ — функция правдоподобия для реализации x.

Def. 18. Оценка максимального правдоподобия Оценкой $\hat{\theta}$ максимального правдоподобия называется статистика, каждая реализция $\theta = \hat{\theta}(\omega)$ которой является точкой максимума функции L правдоподобия при данной реализации $x = X(\omega)$

$$\hat{\theta} : \omega \mapsto \sup_{\theta \in \Theta} f(X(\omega); \theta)$$

Thm. 5. *Уравнения правдоподобия Если $L : \Theta \to [0, 1]$ дифференцируема и при каждой реализации $x \in \mathcal{X}$ супремум достигается в внутренней точке множества Θ, то*

$$\frac{\partial f}{\partial \theta}(x; \theta) = 0$$

или, то же самое:

$$\frac{\partial \ln f}{\partial \theta}(x; \theta) = 0$$

$$\frac{\partial \ln f}{\partial \theta_i}(x; \theta) = 0 \quad \forall x \in \mathcal{X} \quad i = \overline{1, m}$$

Эти уравнения называются уравнениями максимального правдоподобия, а значение оценки максимального правдоподобия (о.м.п.) ищется как её решение при заданной реализации.

0.7 Метод моментов

$X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi) \in \{F_\theta; \theta \in \Theta\}$. $\Theta \subset \mathbb{R}^r$.

Пусть существуют и конечные первые r моментов $\alpha_k(\theta) = \mathbb{E}\xi_\theta^k$. Рассмотрим выборочнные моменты $A_k = \frac{1}{n}\sum_{j=1}^n X_j^k$.

Метод оценивания состоит в решении относительно θ системы равенств

$$a_k(\theta) = A_k(\omega) \qquad k = \overline{1, r}$$

Thm. 6. *Пусть $\tilde{\theta}$ — оценка, полученная методом моментов. $\tilde{\theta}(\omega) = \phi(A_1(\omega), \ldots, A_r(\omega))$. Если ϕ непрерывна и взаимно-однозначна, то так как $\lim_{n\to\infty} \mathbb{P}\{A_k = \alpha_k\} = 1$, то и $\tilde{\theta} \xrightarrow{\mathbb{P}} \theta$*