Список по МС

Сходимость случайных величин

Последовательность случайных величин $X_1, X_2, ...$ сходится к X по вероятности, $X_n \xrightarrow{P} X$, если $\forall \epsilon > 0 : P(|X_n - X| > \epsilon) \to 0$ при $n \to \infty$.

Последовательность случайных величин $X_1, X_2, ...$ сходится слабо к случайной величине X, $X_n \Rightarrow X$, если для каждой точки непрерывности функции распредления $F_X(t)$ имеет место сходимость $F_{X_n}(t) \to F_X(t)$ при $n \to \infty$.

Закон больших чисел

Для любой последовательности $X_1, X_2, ...$ независимых и одинаково распределенных случайных величин с конечным вторым моментом (или конечной дисперсией) $EX_1^2 < \infty$ имеет место сходимость:

$$\frac{X_1 + ... + X_n}{n} \xrightarrow{P} EX_1.$$

Центральная предельная теорема

Для любой последовательности $X_1, X_2, ...$ независимых и одинаково распределенных случайных величин с конечной и ненулевой дисперсией $0 < DX_1 < \infty$ имеет место сходимость:

$$\frac{S_n - nEX_1}{\sqrt{nDX_1}} \Rightarrow N_{0,1},$$

S_n - сумма $X_1 + ... + X_n$.

1. Определение выборки.

Выборка $(X_1, ..., X_n)$ это набор значений случайной величины X, полученный в результате n независимых воспроизведений эксперимента.

2. Определение вариационного ряда, k-ой порядковой статистики.

Если упорядочить элементы выборки по возрастанию, то полученный набор случайных величин называется **вариационным рядом**

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

Случайная величина $X_{(k)}$ называется **k-ой порядковой статистикой.**

3. Определение статистики (оценки).

Статистика - любая функция от выборки.

Оценкой неизвестного параметра θ называется любая функция от выборки $\theta^* = g(X_1, ..., X_n)$ в том или ином смысле приближающая θ.

4. Определение несмещенной оценки.

Оценка θ^* называется **несмещенной**, если $E\theta^* = \theta$.

5. Определение состоятельной оценки.

Оценка θ^* называется **состоятельной**, если $\theta^* \xrightarrow{P} \theta$ при $n \to \infty$.

6. Определение эмпирической функции распределения.

Эмпирической функцией распределения называется

$$F_n^*(y) = \frac{\nu(y)}{n},$$

где $\nu(y)$ - число наблюдений X_i таких, что $X_i < y$.

7. Свойства эмпирической функции распределения.

1. Является случайной величиной.
2. Для любого y: $F^*(y) \to F(y)$ при $n \to \infty$.

8. Как определяется выборочный первый момент (второй, третий и т.д.)?

Выборочный момент порядка k определяется формулой:

$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

9. Как определяется выборочная дисперсия (смещенная и несмещенная)?

Смещенной выборочной дисперсией называется величина:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Несмещенной выборочной дисперсией называется величина:

$$S_0^2 = \frac{n}{n-1}S^2.$$

10. К чему (и как) сходится выборочный первый момент, выборочный второй момент, выборочная дисперсия (смещенная и несмещенная), эмпирическая функция распределения при росте объема выборки?

Первый, второй выборочные моменты сходятся по вероятности к первым и вторым моментам X соответственно.

Выборочная дисперсия (смещенная и несмещенная) также сходится по вероятности к дисперсии X.

Эмпирическая функция распределения также сходится по вероятности к $F(y)$.

11. Чему равно математическое ожидание выборочного первого момента, второго момента, выборочной дисперсии (смещенной и несмещенной), эмпирической функции распределения?

$E\overline{X^k} = \alpha_k$(выборочные моменты - несмещенные оценки теоретических)

$$ES^2 = \frac{n-1}{n}DX_1$$
$$ES_0^2 = DX_1$$
$$E(F_n^*(t)) = F(t)$$

12. Сформулировать теорему Гливенко-Кантелли.

Пусть $X \sim F$, тогда при $n \to \infty$

$$\sup_y |F_n^*(y) - F(y)| \xrightarrow{P} 0.$$

13. Определение ОММ.

Пусть $g(X_i)$ есть некоторая числовая функция $\mathbb{R} \to \mathbb{R}$.

Пусть $m_g(\theta_g^*) = Eg(X_i)$.

Оценкой метода моментов называется такое значение $\theta_g^* = \theta_g^*(\vec{X})$, при котором теоретическое среднее выборки $\vec{g}(X)$ совпадает с выборочным средним:

$$m_g(\theta_g^*) = \overline{g(X)},$$

то есть ОММ является решением уравнения относительно неизвестного θ_g^*.

14. Сформулировать теорему о состоятельности ОММ.

Если $m_g(\theta) = Eg(X_i)$ непрерывна и строго монотонна, то оценка по методу моментов состоятельна.

15. Определение ОМП.

Функцией правдоподобия для выборки $\vec{X}$ называется функция

$$\Pi(\theta) = \Pi(\vec{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta),$$

где $f(X_i, \theta) = \begin{cases} f_{X_i}(t), \text{ при непрерывном распределении} \\ P(X_i = t), \text{ при дискретном распределении} \end{cases}$.

Оценкой максимального правдоподобия называется такое значение параметра $\theta = \hat{\theta}(\vec{X})$, при котором функция правдоподобия принимает наибольшее значение, то есть

$$\Pi(\vec{X}, \hat{\theta}) = \max_{\theta \in \Theta} \Pi(\vec{X}, \theta).$$

16. Найти ОММ для параметров основных распределений (для распределения Бернулли, Пуассона, геометрического, биномиального при известном числе испытаний, равномерного, показательного, нормального распределение при известных и неизвестных параметрах).

1. Бернулли $p^* = \overline{X}$

2. Пуассона $\lambda^* = \overline{X}$
3. Геометрическое $\theta^* = \frac{1}{\overline{X}}$
4. Биномиальное $p^* = \frac{\overline{X}}{m}$ при известном m
5. Равномерное $U_{[0,a]}$ $a^* = 2\overline{X}$
6. Показательное $\lambda^* = \frac{1}{\overline{X}}$
7. Нормальное (при неизвестных)
    1. $a^* = \overline{X}$
    2. ${\sigma^2}^* = S^2$

17. Найти ОМП для параметров основных распределений (для распределения Бернулли, Пуассона, геометрического, биномиального при известном числе испытаний, равномерного, показательного, нормального распределение при известных и неизвестных параметрах).

1. Бернулли $\hat{p} = \overline{X}$
2. Пуассона $\hat{\lambda} = \overline{X}$
3. Геометрическое $\theta^* = \frac{1}{\overline{X}}$(вроде также)
4. Биномиальное $\hat{p} = \frac{\overline{X}}{m}$ при известном m
5. Равномерное $U_{[0,a]}$ $\hat{a} = \max X_1, ..., X_n$
6. Показательное $\hat{\lambda} = \frac{1}{\overline{X}}$
7. Нормальное (при неизвестных)
    1. $\hat{a} = \overline{X}$
    2. $\hat{\sigma}^2 = S^2$

18. Когда одна из оценок не хуже (лучше) в среднеквадратическом смысле другой оценки.

Оценка θ_1^* **лучше**, чем θ_2^*, если при всех значениях θ

$$E(\theta_1^* - \theta)^2 \leq E(\theta_2^* - \theta)^2$$

и хотя бы при одном значении θ неравенство является строгим (иначе **не хуже**).

19. Как сравнить две несмещенные оценки в среднеквадратическом смысле?

Найти их дисперсии и сравнить.

Или найти их вторые моменты и сравнить.

20. Определение эффективной оценки в классе всех несмещенных оценок.

Оценка с минимальной дисперсией среди всех несмещенных называется **эффективной**.

21. Определение доверительного интервала (точного, асимптотического).

Доверительным интервалом уровня $1-\epsilon$ для неизвестного параметра θ называется интервал $(A(\vec{X}), B(\vec{X}))$, такой, что

$$P(A < \theta < B) \geq 1-\epsilon$$

Доверительный интервал называют **асимптотическим**, если

$$\lim_{n\to\infty} P(A < \theta < B) \geq 1-\epsilon$$

Доверительный интервал называют **точным**, если знак неравенства можно заменить на знак равенства (в обоих случаях).

22. Схема построения ДИ (точного, асимптотического).

1. Строим некую функцию $G(\vec{X}, \theta) \sim H_n$, где H_n известно.
2. $P(t_1 < G < t_2) = 1-\epsilon$, t_1, t_2 - квантили.
3. Решаем двойное неравенство в вероятности для θ.

Для асимптотического - G должна стремиться к $\eta \sim H_n$.

23. Определение распределения Хи-квадрат, Стьюдента, Фишера.

Говорят, что с. в. Z имеет **распределение χ^2 с m степенями свободы**, если

$Z = \xi_1^2 + ... + \xi_m^2$, где $\xi_i \sim N_{0,1}$, независимы, $i = 1, ..., m$.

Говорят, что с. в. Y имеет **распределение Стьюдента с m степенями свободы**, если

$Y = \frac{\xi_0}{\sqrt{\frac{\xi_1^2 + ... + \xi_m^2}{m}}}$, где $\xi_i \sim N_{0,1}$, независимы, $i = 0, ..., m$. (T_m)

Говорят, что с. в. Y имеет **распределение Фишера с m и k степенями свободы**, если

$Y = \frac{Z_1/m}{Z_2/k}$, где $Z_1 \sim \chi_m^2$ и $Z_2 \sim \chi_k^2$, независимы. ($F_{m,k}$)

24. Сформулировать лемму Фишера, следствия из леммы Фишера.

Лемма Фишера. Пусть случайные величины $\vec{X} = (X_1, ... X_n)$ независимы, $X_i \sim N_{0,1}$, $i = 1, ..., n$, и $\vec{Y} = A\vec{X}$, A - ортогональная матрица. Тогда для любого $r = 1, ..., n-1$

$$X_1^2 + ... + X_n^2 - Y_1^2 - ... - Y_r^2 \sim \chi_{n-r}^2$$

и эта случайная величина не зависит от $Y_1, ..., Y_r$.

Следствия. $\vec{X} \sim N_{a,\sigma^2}$, тогда:

1. $\frac{\overline{X} - a}{\sigma}\sqrt{n} \sim N_{0,1}$
2. $\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2$
3. $\overline{X}$ и S^2 независимы.

25. Сформулировать теорему Фишера (4 утверждения).

Следствие основного следствия леммы Фишера.

1) $\sqrt{n}\dfrac{\overline{X}-a}{\sigma}$ имеет стандартное нормальное распределение (для a при σ известном);

2) $\sum\limits_{i=1}^{n}\left(\dfrac{X_i-a}{\sigma}\right)^2$ имеет распределение H_n (для σ^2 при a известном);

3) $\sum\limits_{i=1}^{n}\left(\dfrac{X_i-\overline{X}}{\sigma}\right)^2 = \dfrac{(n-1)S_0^2}{\sigma^2}$ имеет распределение H_{n-1} (для σ^2 при a неизвестном);

4) $\sqrt{n}\dfrac{\overline{X}-a}{\sqrt{S_0^2}} = \sqrt{n}\dfrac{\overline{X}-a}{S_0}$ имеет распределение T_{n-1} (для a при σ неизвестном).

1. $\sum\limits_{i=1}^{n}(\frac{X_i-a}{\sigma})^2 = \frac{nS_1^2}{\sigma^2} \sim \chi_n^2$
2. $\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2$
3. $\sqrt{n}(\frac{\overline{X}-a}{\sigma}) \sim N_{0,1}$
4. $\sqrt{n}(\frac{\overline{X}-a}{S_0}) \sim T_{n-1}$

26. Выписать доверительные интервалы для параметров нормального распределения при известных и неизвестных параметрах математического ожидания и дисперсии (таких ДИ четыре штуки).

Для a, σ^2 известно:

$P(\overline{X} - \tau_{1-\epsilon/2}\frac{\sigma}{\sqrt{n}} < a < \overline{X} + \tau_{1-\epsilon/2}\frac{\sigma}{\sqrt{n}}) = 1-\epsilon$, квантиль $N_{0,1}$

Для a, σ^2 неизвестно:

$P(\overline{X} - t\frac{S_0}{\sqrt{n}} < a < \overline{X} + t\frac{S_0}{\sqrt{n}}) = 1-\epsilon$, $t = -\tau_{\epsilon/2}$, квантиль T_{n-1}

Для σ^2, a известно

$P\left(\frac{\sum_{i=1}^{n}(X_i-a)^2}{q_2} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-a)^2}{q_1}\right) = 1-\epsilon$, $\chi_n^2(q_1) = \epsilon/2, \quad \chi_n^2(q_2) = 1-\epsilon/2$

Для σ^2, a неизвестно

$P\left(\frac{nS^2}{q_2} < \sigma^2 < \frac{nS^2}{q_1}\right) = 1-\epsilon$, $\chi_{n-1}^2(q_1) = \epsilon/2, \quad \chi_{n-1}^2(q_2) = 1-\epsilon/2$

27. Определение гипотезы (простой гипотезы, сложной гипотезы).

Гипотеза - любое суждение о неизвестном распределении.

Гипотеза называется **простой**, если она однозначно определяет распределение выборки. Иначе - **сложной**.

28. Определение критерия.

$H_1, ..., H_m$ - гипотезы.

Критерием называется отображение $\delta : \mathbb{R}^n \to 1, ..., m$. Это правило, по которому выбирается одна из гипотез.

29. Определение вероятностей ошибок i-го рода.

Вероятность **ошибки первого рода** - вероятность отвергнуть верную основную гипотезу.

Вероятность **ошибки второго рода** - вероятность принять неверную основную гипотезу.

30. Определение мощности, состоятельности в критериях согласия.

Если β_2 - вероятность ошибки 2 рода, то число $1 - \beta_2$ называется **мощностью** критерия.

Если мощность критерия стремится к 1, то есть вероятность ошибки 2 рода стремится к 0, то критерий называется **состоятельным**.

Вероятность ошибки первого рода также называется **размером** критерия.

31. Общий вид критериев согласия. Какие два условия накладываются на статистику в критериях согласия.

Пусть $\vec{X} \sim F$

$H_0 = \{F = F_0\}$

$H_a = \{F \neq F_0\}$

Нужно придумать $d(F_0, F_n^*)$, такое, что:

1. При верной $H_0 : d(F_0, F_n^*) \Rightarrow \eta \sim G$, известное
2. При верной $H_a : d(F_0, F_n^*) \to \infty$

Тогда **критерием согласия** называется критерий:

$$\delta = \begin{cases} 0, & d(F_0, F_n^*) < c \\ 1, & d(F_0, F_n^*) \geq c \end{cases}$$

где $G(c) = 1 - \epsilon$.

32. Сформулировать теорему Колмогорова, теорему Пирсона.

Теорема Колмогорова. Пусть $X \sim F$, F - непрерывна. Если

$$d = \sup_y |F_n^*(y) - F(y)|$$

то для любого $y > 0$ при $n \to \infty$

$$P(\sqrt{n}d < y) = \mathcal{K}(y),$$

$\mathcal{K}(y)$ - функция Колмогорова, табулированная.

ИЛИ

$$\sqrt{n} \sup_y |F_n^*(y) - F(y)| \Rightarrow \mathcal{K}$$

Теорема Пирсона.

$X \sim F$, проверяются гипотезы как обычно.

Разобьем область возможных значений X_1 на k непересkающихся промежутков Δ_j.

$p_j = P_{H_0}(X_1 \in \Delta_j)$,

ν_j - число наблюдений, попавших j-ый промежуток,

$\Psi_n = \sum\limits_{i=1}^{k} \frac{(\nu_j - np_j)^2}{np_j}$

Если $0 < p_i < 1$ при всех $i = 1, ..., k$, то для любого $y > 0$, при верной H_0

$$\Psi_n \Rightarrow \eta \sim \chi^2_{k-1}.$$

33. Выписать критерий Колмогорова.

$$\Psi_n = \sqrt{n}\sup_y |F_n^*(y) - F(y)|$$

$$\delta = \begin{cases} 0, & \Psi_n < c \\ 1, & \Psi_n \geq c \end{cases}$$

где $\mathcal{K}(c) = 1 - \epsilon$.

34. Выписать критерий Хи-квадрат (Пирсона).

$$\Psi_n = \sum_{i=1}^{k} \frac{(\nu_j - np_j)^2}{np_j}$$

$$\delta = \begin{cases} 0, & \Psi_n < c \\ 1, & \Psi_n \geq c \end{cases}$$

где $\chi^2_{k-1}(c) = 1 - \epsilon$.

35. Сформулировать теорему Стьюдента (для критерия Стьюдента о равенстве средних), теорему Фишера (для построения критерия Фишера о равенстве дисперсий).

Теорема Стьюдента. (если средние совпадают) n, m - размеры независимых (нормальных) выборок X, Y соответственно.

$$\frac{\overline{X} - \overline{Y}}{\sqrt{1/n + 1/m}\sqrt{\frac{nS_X^2 + mS_Y^2}{n+m-2}}} \sim T_{n+m-2}$$

Теорема Фишера. n, m - размеры независимых нормальных выборок X, Y соответственно, тогда (если дисперсии совпадают)

$$\frac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} \sim F_{n-1,m-1}$$

36. Выписать критерий Стьюдента.

τ - квантиль данного распределения.

$$d = \frac{\overline{X} - \overline{Y}}{\sqrt{1/n + 1/m}\sqrt{\frac{nS_X^2 + mS_Y^2}{n+m-2}}} \sim T_{n+m-2}$$

$$\delta = \begin{cases} 0, & |d| < \tau_{1-\epsilon/2} \\ 1, & |d| \geq \tau_{1-\epsilon/2} \end{cases}$$

37. Выписать критерий Фишера.

f - квантиль данного распределения.

$$d = \frac{S_0^2(X)}{S_0^2(Y)} \sim F_{n-1,m-1}$$

$$\delta = \begin{cases} 0, & f_{\epsilon/2} < d < f_{1-\epsilon/2} \\ 1, & \text{иначе} \end{cases}$$

38. Дать определение реально достигнутого уровня значимости.

Реально достигнутым уровнем значимости называется число $\epsilon^* = P_{H_0}(d(F_n^*, F_0) > \tilde{d})$.

$\tilde{d}$ - полученное значение статистики.

Если РДУЗ мал, то отвергаем H_0.

Если РДУЗ близок к 1, то принимаем H_0.