Дополнительные главы математической статистики (непараметрическая статистика)

Вопросы и задачи к курсу

осенний семестр 2013г.

Задачи о ранговых статистиках.

1. Определите статистику Уилкоксона для двух выборок . Покажите, что для однородных выборок эта статистика распределена свободно. Для этого случая найдите её математическое ожидание и дисперсию. Укажите связь статистик Уилкоксона и Манна-Уитни.

2. Определите статистику Манна-Уитни для двух выборок . Покажите, что для однородных выборок эта статистика распределена свободно. Для этого случая найдите её математическое ожидание и дисперсию. Укажите связь статистик Уилкоксона и Манна-Уитни.

3. Определите двувыборочную Ю-статистику. Вычислите её математическое ожидание и дисперсию. Докажите закон больших чисел для этих статистик.

4. Докажите центральную предельную теорему для двувыборочных Ю-статистик.

5*. Покажите, что статистики Уилкоксона и Манна-Уитни распределены асимптотически нормально.

6*. Рассмотрим две выборки, отличающиеся сдвигом. С помощью статистики Уилкоксона для этого сдвига укажите оценку и доверительные интервалы.

7*. Покажите, что выборочные квантили распределены асимптотически нормально и укажите их асимптотические средние и дисперсии. Особо рассмотрите выборочную медиану.

8*. Определите медиану Ходжеса-Лемана для двух выборок. Покажите, что она распределена асимптотически нормально и укажите её асимптотические среднее и дисперсию.

9. Определите одновыборочные Ю-статистики. Для одновыборочной Ю-статистики вычислите математическое ожидание, дисперсию и докажите закон больших чисел.

10*. Докажите центральную предельную теорему для одновыборочных Ю-статистик.

11. Определите статистику знаковых рангов Уилкоксона для парных наблюдений. Вычислите её математическое ожидание и дисперсию при гипотезе.

12*. Покажите, как статистика знаковых рангов Уилкоксона выражается через Ю-статистики. На этом основании докажите её асимптотическую нормальность.

13. Пусть $R_1, R_2, ..., R_n$ суть ранги n независимых одинаково и непрерывно распределенных случайных величин. Вычислите ER_1, DR_1, ER_1R_2, Коэффициент корреляции между R_1 и R_2.

Задачи о статистике Колмогорова-Смирнова.

14. Укажите представление винеровского процесса на отрезке (0,1) в виде ряда.

15. Укажите представление брауновского моста на отрезке (0,1) в виде ряда.

16*. Для предельного распределения статистики омега-квадрат укажите представление в виде ряда.

17*. По определению, броуновский мост $\beta(t),\ 0 \le t \le 1$ - это гауссовский случайный процесс с непрерывными траекториями, для которого $E\beta(t) = 0,\ E\beta(s)\beta(t) = \min(s,t) - st$. Покажите, что распределение процесса $\beta(t)$ совпадает с условным распределением винеровского процесса $W(t),\ 0 \le t \le 1$, при условии $W(1) = 0$.

18. Покажите, что случайный процесс $W(t) - tW(1),\ 0 \le t \le 1$, по распределению совпадает с $\beta(t)$.

19*. Для выборки объема n из распределения с непрерывной функцией $F(\cdot)$ определим случайный процесс $\beta_n(t)$, положив

$$\beta_n(t) = \sqrt(n)\,[F_n(x) - F(x)], \qquad \text{где } t = F(x).$$

Покажите, что конечномерные распределения случайного процесса $\beta_n(t)$ сходятся к соответствующим распределениям процесса $\beta(t)$, при $n \to \infty$.

20*. Докажите центральную предельную теорему для полиномиальных распределений (эта теорема может быть полезна при решении предыдущей задачи).

21*. Рассмотрим проверку гипотезы об однородности двух выборок равных объемов с помощью статистики Колмогорова-Смирнова. Вычислите распределения этих статистик (при гипотезе), сведя дело к случайным блужданиям.

22*. (Продолжение 5). Переходом к пределу (при неограниченном увеличении объема выборок) получите известную формулу Смирнова (для броуновского моста).

Задачи о порядковых статистиках.

Пусть $x_1, x_2, ..., x_n$ - выборка из распределения, имеющего плотность $f(\cdot)$. Упорядочив элементы выборки по возрастанию, получим т.н. вариационный ряд $x_{(1)}, x_{(2)}, ..., x_{(n)}$. Члены этого ряда называют порядковыми статистиками. Это случайные величины, и они тоже имеют плотности распределения.

23. Укажите

a. Совместную плотность n порядковых статистик;
b. Плотность $x_{(1)}$, плотность $x_{(n)}$;
c. Плотность $x_{(k)}$ для данного $k = 1, ..., n$.
d. Совместную плотность $x_{(k)},\ x_{(l)}$ для $k < l$.
e. Плотность медианы выборки - для простоты при $n = 2m - 1$.

24. Здесь рассматривается выборка из распределения, равномерного на $(0;1)$. Вычислите:

a. $E\xi_{(k)}$ для $k = \overline{1,n}$;
b. Асимптотическое распределение $[tn]$-ой порядковой статистики для фиксированного $t,\ 0 < t < 1$.

25. Рассмотрим ранги $R_1, R_2, ..., R_n$ независимых одинаково и непрерывно распределенных случайных величин.

Укажите распределение каждого $R_1, ..., R_n$, каждой пары (R_k, R_l), где $k \neq l$; Каждого набора $R_{i_1}, ..., R_{i_k}$. Вычислите $ER_1,\ ER_1R_2,\ DR_1,\ cov(R_1, R_2)$.

26*. Ранги наблюдений $x_1, x_2, ..., x_n$ инвариантны при монотонно возрастающих преобразованиях координат. Более того, ранги - это максимальный инвариант. Статистики Колмогорова-Смирнова при упомянутых преобразованиях инвариантны. Поэтому они должны быть функциями рангов. Как статистики К.-С. выражаются через ранги наблюдений.

27. Пусти $a(k) = -E\frac{f'((\xi_k))}{f((\xi_k))},\ k = \overline{1,N}$ - система меток. Покажите, что $\sum_{k=1}^{N} a(k) = 0$.

28*. Дайте математические обоснования для предложений §5.1 из книги М. Холландер, Д.А. Вулф "Непараметрические методы статистики"/Пер. с англ. - М.:Финансы и Статистика, 1983г. (Там описан ранговый тест Ансари-Бредли (Ausari,Bradley, 1960))

29*. Найдите наилучший ранговый критерий в задаче о сравнении двух показательных выборок, различающихся масштабом. Вычислите метки для данного критерия.

30*. Даны две независимые выборки из двусторонних показательных распределений, отличающихся сдвигом. Укажите наилучший ранговый критерий (и его метку) для проверки гипотезы однородности.

Задачи о ранговых оценках.

31*. Пусть $x_1, x_2, .., x_n$ - независимые реализации случайной величины x, распределенной симметрично относительно некоторой точки θ. Рассмотрим выборочную псевдомедиану x, положив

$$\hat{\theta}_n = \text{медиана}\{\frac{1}{2}(x_i + x_j)|1 \leq i < j \leq n\}$$

a) Покажите, что $\hat{\theta}_n$ сходится к θ, при $n \to \infty$.

b) Покажите, что, при определенных условиях, $\hat{\theta}_n$ распределена ассимптотически нормально.

c) Найдите ассимптотическое среднее и ассимптотическую дисперсию $\hat{\theta}_n$.

32*. Вычислите математическое ожидание и дисперсию коэффициенты ранговой корреляции

a) Спирмена, в случая независимых признаков;

b) Кендалла, в случае независимых признаков и в общем случае.

33*. Докажите, что коэффициент ранговой корреляции Кендалла распределен ассимптотически нормально.

34. По наблюдениям линейной регрессии y по x

$$y_k = a + bx_k + \epsilon_k, \qquad k = \overline{1, n},$$

когда случайные ошибки независимы и одинаково распределены, укажите для коэффициента наклона статистическую оценку и доверительные интервалы, используя коэффициент ранговой корреляции.

Литература

1. Т.Хеттманспергер. Статистические выводы, основанные на рангах./ Пер. С англ.;– М.: Финансы и статистика, 1987. – 334с.

2. Я.Гаек,З.Шидак. Теория ранговых критериев./ Пер. С англ.; – М.: "Наука 1971. – 376с.

3. М.В.Болдин, Г.И.Симонова, Ю.Н.Тюрин. Знаковый статистический анализ линейных моделей. – М.: Наука, 1997. – 288с.

4. М.Холлендер, Д.Вулф. Непараметрические методы статистики. / Пер. С англ. – М.: Финансы и статистика, 1983. – 518с.