

Математическое моделирование и обработка данных (4/7),  
Ермаков М.С. (2012/2013 уч.год)

1. Задачи математической статистики. Параметрические, семипараметрические, непараметрические статистические выводы. Описательная статистика. Теория статистических выводов. Типы данных: количественные и качественные; первичные, вторичные (степень их достоверности). Случайная и систематическая ошибки. Примеры.
2. Выборка. Статистика. Частота. Относительная частота. Связь относительной частоты события и вероятности события.
3. Эмпирическая функция распределения и гистограмма. Их связь с теоретической функцией распределения и плотностью распределения. Уметь их строить.
4. Основные числовые характеристики выборки и общий принцип построения их оценок. Выборочные среднее и дисперсия, начальный и центральный моменты.
5. Оценки параметров положения. Медиана, мода, выборочное среднее.
6. Оценки параметров рассеяния. Выборочная дисперсия, стандартное отклонение, абсолютное отклонение, квантили, квартили, персентили, интерквартильный размах, размах.
7. Оценки параметров формы плотности. Асимметрия и эксцесс.
8. Роль нормального распределения в статистике.
9. Теорема об асимптотической нормальности функций от асимптотически нормальных случайных величин. Ее следствия.
10. Теория статистического оценивания. Общая постановка задачи. Определения статистики, статистической оценки, квадратичной функции риска и ее связь с дисперсией оценки. Примеры статистических оценок.
11. Доверительное оценивание (зачем оно нужно?). Доверительный интервал, уровень доверия.
12. Доверительный интервал для математического ожидания нормального распределения (дисперсия известна и неизвестна). Доверительный интервал для дисперсии нормального распределения. Доверительный интервал для биномиального распределения.
13. Построение доверительного интервала на основе асимптотически нормальных статистических оценок. Многомерное нормальное распределение. Доверительный эллипсоид.
14. Многомерное нормальное распределение для произвольной ковариационной матрицы. Его плотность и график. Болезнь размерности.
15. Арифметика нормального распределения. Независимость выборочных средних и дисперсии.
16. Распределения, связанные с нормальным: хи-квадрат, Стьюдента, Фишера. Где каждое из них применяется.
17. Требования к статистическим оценкам. Состоятельность, несмещенность, асимптотическая нормальность.
18. Несмещенные оценки. Неравенство Рао-Крамера. Информационное количество Фишера. Эффективность выборочного среднего в нормальной модели.
19. Асимптотическая нормальность. Стандартная ошибка. Асимптотическая эффективность. Нижняя граница в асимптотически эффективном оценивании.
20. Метод моментов. Обобщенный метод моментов. Метод наименьших квадратов и его геометрическая интерпретация. Метод хи-квадрат.
21. Байесовский подход к теории оценивания. Априорная и апостериорная плотности распределения. Апостериорный байесовский риск.
22. Три интерпретации байесовского подхода. Примеры байесовских оценок. Оценки максимума апостериорной вероятности.
23. Функция правдоподобия. Оценки максимума правдоподобия (ОМП).
24. Асимптотическая нормальность и эффективность ОМП.
25. Асимптотическое поведение отношения правдоподобия.
26. Построение по любой состоятельной оценке асимптотически эффективной и асимптотически нормальной оценки.
27. Семипараметрическое оценивание. Функция влияния и ее содержательный смысл. Производные по Гато, Фреше и Адамару. Связь с робастностью. Асимптотическая нормальность.
28. Вычисление производных по Гато статистических оценок.
29. М-оценки. Их асимптотическая нормальность.
30. Минимаксный подход к задачам статистического оценивания. Связь минимаксного и байесовского рисков. Минимаксность выборочного среднего.

31. Проверка статистических гипотез. Простые и сложные гипотезы. Статистический критерий. Область принятия гипотезы. Критическая область. Тестовая статистика. Вероятности ошибок первого и второго рода. Статистические аспекты выбора гипотезы и альтернативы. Мощность критерия. Р-значение.

32. Лемма Неймана-Пирсона. Наиболее мощные критерии. Критерий отношения правдоподобия. Замечание к лемме Неймана Пирсона.

33. Равномерно наиболее мощные (РНМ) и равномерно наиболее мощные несмещенные (РНМН) критерии в нормальной модели. Примеры. Содержательный смысл понятия несмещенности.

34. Связь доверительного оценивания и проверки гипотез.

35. Асимптотически эффективные критерии. Асимптотическая эффективность по Питману. Проверка гипотез на основе асимптотически нормальных оценок.

36. Байесовские критерии проверки сложных гипотез. Критерий максимума отношения правдоподобия для проверки сложных гипотез.

37. Корреляционный и регрессионный анализ. Выборочная ковариация и корреляция. Уметь считать.

38. Функция регрессии. Многомерная линейная регрессия. Метод наименьших квадратов. Матричная запись метода наименьших квадратов. Геометрическая интерпретация метода наименьших квадратов.

39. Метод наименьших квадратов для больших размерностей. Метод регуляризации, штрафная функция.

40. Решение интегрального уравнения методом регуляризации.

41. Многомерные гауссовские наблюдения и модель линейной регрессии.

42. Простая линейная регрессия. Свойства оценок МНК простой линейной регрессии.

43. Коэффициент детерминации, его содержательный смысл и связь с выборочным коэффициентом корреляции. Основное тождество " $SST = SSE + SSR$ " его содержательный смысл.

44. Нелинейная регрессия. МНК. Его геометрическая интерпретация. Асимптотическая нормальность оценок МНК.

45. Логистическая регрессия. Оценка максимума правдоподобия (уметь ее написать; уравнения, задающие оценки, запоминать необязательно).

46. Jackknife. Оценка дисперсии.

47. Бутстрап. Построение несмещенных оценок.

48. Бутстрап параметрический и непараметрический. Оценка дисперсии и доверительного интервала. Состоятельность бутстрапа. Теорема Гливленко-Кантели.

49. Непараметрическое оценивание. Проекционные оценки плотности, их состоятельность в  $L_2$ .

50. Sequence model как модель оценки плотности.

51. Задача оценивания функции регрессии, sequence model для нее, ее связь с оцениванием сигнала в гауссовском белом шуме.

52. Критерии существования состоятельных оценок в  $L_2$ . Критерий компактности в  $L_2$ .

53. Примеры компактов в  $L_2$ . Шары в пространстве Соболева.

54. Вейвлеты и болезнь размерности. Вейвлеты Хаара. Зачем они нужны. Father and mother вейвлеты.

55. Soft and hard thresholding. Выбор thresholding параметра.

57. Ядерные оценки плотности, их состоятельность.

58. Оптимальный выбор ширины окна в зависимости от гладкости.

59. Непараметрическое оценивание регрессии.

Очень сложные формулы, распределения, вид доверительных интервалов, константы разрешается точно не помнить — однако это на мое усмотрение. На каждый вопрос программы иметь краткий ответ без шпаргалки, чтобы ответить перед моими глазами. Обязательно прочтите в конце подготовки программу и поймите, что вы можете ответить, а что нет.

При решении задачи, если не уверены в правильности перевода, пишите его и подходите ко мне. Когда задача решена, открываете тетрадку, проверяете по основным ошибкам, упомянутым в ней. При решении задачи на листке обязательно указывается фамилия, постановка задачи (например вид гипотезы и альтернативы), а также основная формула, по которой решается задача. На решение задачи даётся 35 минут. На вопрос 45 минут. Если вы подошли ко мне и задача неправильна, минус балл.

Кто не получит допуск — 5 доп.задач на экзамене + за каждое неверное решение ещё две.

Не пуха, ни пера — однако, ошипанными не приходиться.

М.Ермаков