СОДЕРЖАНИЕ

Программа учебного курса «Математическая статистика»

Стандартные распределения в статистическом анализе данных. Распределение хи-квадрат (Пирсона). Случайная величина хи-квадрат как сумма квадратов независимых стандартных нормальных случайных величин. Распределения Фишера–Снедекора и Стьюдента. Распределение Дирихле и бета-распределение.

Точечное оценивание параметра закона распределения. Состоятельность, несмещенность (асимптотическая несмещенность) и оптимальность точечной оценки. Неравенство Рао–Крамера (скалярный и векторный случаи). Эффективность (асимптотическая эффективность) точечной оценки.

Достаточность статистики относительно параметра, критерий факторизации. Теорема Блекуэла–Рао о возможности улучшения оценок при наличии достаточных статистик.

Полные достаточные статистики. Теорема об оптимальности полных достаточных статистик.

Метод моментов. Функция правдоподобия. Метод наибольшего правдоподобия и свойства получаемой оценки. Теорема о состоятельности, асимптотической несмещенности, нормальности и эффективности оценки наибольшего правдоподобия.

Интервальное оценивание параметра закона распределения, доверительный интервал. Свойства статистик, используемых для интервального оценивания. Построение доверительных интервалов параметров нормального распределения. Интервальное оценивание при больших выборках. Проверка гипотез о равенстве математических ожиданий и равенстве дисперсий двух нормальных случайных величин с использованием доверительных интервалов.

Метод наименьших квадратов. Точечное оценивание векторного параметра. Система нормальных уравнений. Свойства оценки метода наименьших квадратов, теорема Гаусса–Маркова. Интервальное оценивание по методу наименьших квадратов. Нормальная регрессия. Теорема об оптимальности оценки метода наименьших квадратов в случае нормальной регрессии.

Эмпирическая функция распределения. Статистическая гипотеза, выборка, критическая область гипотезы, уровень значимости. Теоремы о свойствах критериев согласия Колмогорова и Пирсона (хи-квадрат). Теорема Крамера (параметрический хи-квадрат).

Применение критериев для проверки согласия результатов опыта с теоретическими гипотезами о виде функции распределения, об однородности выборок, о независимости случайных величин. Использование статистических таблиц.

Порядковые статистики и их законы распределения. Доли и блоки выборки. Статистическая эквивалентность блоков выборки. Распределение Пуассона–Дирихле. Задачи непараметрического оценивания.

Выбор статистических гипотез. Простые и сложные статистические гипотезы. Статистическое решение и решающее правило. Рандомизированные решающие правила. Ошибки первого и второго рода, мощность статистического критерия. Наиболее мощный и равномерно наиболее мощный критерий.

Критерий Неймана–Пирсона для двух простых гипотез. Решающее правило, оптимальное по критерию Неймана–Пирсона. Функция отношения правдоподобия. Лемма Неймана–Пирсона о построении решающего правила.

Критерий максимума апостериорной вероятности.

Критерий минимума среднего риска (Байеса) в случае простых гипотез. Функция штрафа, функция риска, средний риск. Решающее правило в случае двух простых гипотез, в случае произвольного конечного числа простых гипотез.

Критерий Байеса в случае сложных гипотез. Решающее правило при известной функции распределения случайного параметра. Минимаксная процедура для случая неизвестного закона распределения случайного параметра.

Связь критерия Байеса с критерием максимума апостериорной вероятности, минимаксным критерием, критерием Неймана–Пирсона.

Последовательный критерий отношения вероятностей (критерий Вальда) и построение последовательной процедуры выбора при двух простых гипотезах. Теорема об оптимальности критерия отношения правдоподобия. Вид решающего правила на произвольном шаге процедуры и его интерпретация. Теорема о завершении процедуры за конечное число шагов с вероятностью единица. Выбор параметров решающего правила при заданных величинах вероятностей ошибок первого и второго рода.

Задачи
по курсу «Математическая статистика»

1. Пусть $F_n(x)$ – эмпирическая функция распределения, получаемая по простой выборке $X_1, \ldots, X_n$ случайной величины X, обладающей функцией распределения $F(x)$. Оценить при больших n вероятность события $\{|F_n(x^*) - F(x^*)| \le \frac{t}{\sqrt{n}}\}$ (для заданного t и при $0 < F(x^*) < 1$).

2. Пусть X – случайная величина с заданной функцией распределения $F(x)$. Найти совместную функцию распределения порядковых статистик $X_{(r)}$ и $X_{(s)}$ ($1 \le r < s \le n$, n – объем выборки).

3. Пусть случайная величина X имеет равномерное распределение на отрезке $[a, b]$. Найти совместное распределение минимального $X_{(1)}$ и максимального $X_{(n)}$ элементов ее простой выборки $X_1, \ldots, X_n$. Вычислить их математические ожидания, дисперсии и коэффициент корреляции.

4. Предполагается выполнить $n + 1$ независимых измерений случайной величины X, имеющей непрерывную функцию распределения $F(x)$. Найти: *а*) априорную вероятность того, что значение X_{n+1}, полученное в $(n + 1)$-м измерении, окажется больше, чем k-е по величине значение X, полученное в предшествующих n измерениях; *б*) априорную вероятность того, что значение X_{n+1} окажется в k-м блоке выборки, т.е. вероятность $P\{X_{(k)} < X_{n+1} < X_{(k+1)}\}$. Зависит ли она от номера блока?

5. Пусть $X_1, ..., X_{2n-1}$ – простая выборка случайной величины, имеющей непрерывную функцию распределения $F(x)$, а $W_1, ..., W_{2n}$ – ее доли ($W_i = F(x_{(i)}) - F(x_{(i-1)})$). Найти распределение вероятностей суммы $S = \sum_{k=1}^{n} W_{2k}$ (суммы «четных» долей).

6. Пусть числа $U_2 > U_1 > 0$ такие, что величина $U_2 - U_1$ достигает минимума при условии $P(X_{10}^2 \in [U_1, U_2]) = \alpha$. Дока-

зать, что U_1 и U_2 являются корнями уравнения $\frac{u^4 e^{-u/2}}{768} = \lambda$ для некоторого $\lambda > 0$.

7. При $N = 4040$ бросаниях монеты получено $N_1 = 2048$ выпадений "герба" и $N_2 = 1992$ выпадений "решетки". Согласуются ли результаты с гипотезой о "симметричности" монеты при уровне значимости $\alpha = 0{,}05$?

8. При 72 бросаниях игральной кости грани "1", "2", "3", "4", "5", "6" выпали 9, 20, 14, 8, 11, 10 раз соответственно. Можно ли считать игральную кость "симметричной" при уровне значимости $\alpha = 0{,}01$?

9. Цифры 0, 1, 2, ... , 9 среди 800 первых десятичных знаков числа π появляются 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 раз соответственно. Проверить гипотезу о согласии данных с законом равномерного распределения.

10. При эпидемии гриппа из 200 контролируемых людей однократное заболевание наблюдалось у 181 человека, а дважды болели гриппом 9 человек. Правдоподобна ли гипотеза о том, что в течение эпидемии гриппа число заболеваний отдельного человека представляет собой случайную величину, подчиняющуюся биномиальному распределению с числом испытаний $n = 2$?

11. Произведено измерение размеров деталей в двух партиях деталей по 100 деталей в каждой партии. В первой партии оказалось 25 деталей с заниженным размером, 50 деталей с точным размером, 25 деталей с завышенным размером, а во второй партии аналогичные числа оказались равны 52, 41, 7 соответственно. Проверить гипотезу о независимости номера партии деталей и размера детали.

12. При снятии показаний измерительного прибора десятые доли деления шкалы прибора оцениваются "на глаз" наблюдателем. Количества цифр 0, 1, 2, ..., 9, записанных наблюдателем в качестве десятых долей при 100 независимых измерениях, равны 5, 8, 6, 12, 14, 18, 11, 6, 13, 7 соответственно. Проверить гипотезы о согласии данных с законом равномерного распреде-

ления и с законом нормального распределения. Для ответа на вопрос можно сравнить значения $p-\text{value}$ для обеих гипотез.

Комментарии:

Пусть Ω_α – критическая область для критерия с уровнем значимости α. Используемая статистика критерия $T(X)$. Тогда

$$p-\text{value}=\inf\{\alpha:\quad T(X)\in\Omega_\alpha\}.$$

13. Пусть T_n – состоятельная оценка для параметра θ, а φ(x) – непрерывная функция. Доказать, что φ(T_n) – состоятельная оценка для φ(θ).

14. Пусть $(X_1,Y_1,)$, (X_2,Y_2), ..., (X_n,Y_n) – простая выборка из генеральной совокупности (X, Y). Показать, что величина

$$\hat{R}_{XY}=\left[\sum_{r=1}^{n}(X_r-\bar{X})(Y_r-\bar{Y})\right]/(n-1)$$

является несмещенной и состоятельной оценкой корреляционного момента $R_{XY}=\operatorname{cov}(X,Y)=M(\overset{0}{X}\overset{0}{Y})$.

15. Используя таблицу случайных чисел, получить реализацию выборки $x_1, x_2, ..., x_n$ из равномерно распределенной на отрезке [0,1] генеральной совокупности X (значения x_i взять с двумя десятичными знаками, n = 50). Найти:

а) вариационный ряд $x_{(1)}\le x_{(2)}\le\ldots\le x_{(n)}$;

б) эмпирическую функцию распределения (построить ее график и график теоретической функции распределения);

в) $\bar{x}=(\sum_{i=1}^{n}x_i)/n$ (сравнить с MX);

г) $S^2=\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]/(n-1)$ (сравнить с DX).

Используя критерий χ^2, проверить гипотезу о соответствии полученной реализации выборки равномерному распределению на отрезке [0, 1] при уровне значимости $\alpha = 0{,}05$.

16. Используя таблицу случайных чисел, получить реализацию выборки $x_1, x_2, ..., x_n$ из нормально распределенной, генеральной совокупности $X : X \in \mathsf{N}\left(\frac{1}{2}, 1\right)$ (значения x_i взять с двумя десятичными знаками, $n = 50$). Найти вариационный ряд, эмпирическую функцию распределения, вычислить $\overline{x}$, S^2 (см. п п. *в*, *г* задачи № 15). Используя критерий χ^2, проверить гипотезу о соответствии полученной реализации выборки нормальному распределению (при неизвестных математическом ожидании и дисперсией) при уровне значимости $\alpha = 0{,}05$.

17. Пусть $X_i \in \mathsf{N}(m, \sigma^2)$, $i = \overline{1, n}$, где m и σ^2 неизвестны. Используя метод моментов, построить оценку параметра m по результатам измерений $V_i = \exp(X_i)$ $i = \overline{1, n}$.

18. Построить состоятельные оценки параметров m и p по результатам измерения k независимых случайных величин, каждая из которых с вероятностью p подчиняется распределению $\mathsf{N}(0, 1)$, а с вероятностью $1–p$ – распределению $\mathsf{N}(m, 1)$, где $-\infty < m < \infty$, $0 \le p \le 1$ (рекомендуется воспользоваться методом моментов).

19. При измерении длины стержня, истинная длина которого равна $l > 0$ (и неизвестна), ошибка измерения имеет распределение $\mathsf{N}(0, k\,l)$, где k – известное число. Найти оценку наибольшего правдоподобия для параметра l, построенную на основании независимых измерений $X_1, X_2, ..., X_n$ длины стержня.

20. Найти оценки наибольшего правдоподобия и эффективные оценки (если они существуют):

а) параметра λ в пуассоновском распределении;

б) параметра μ в показательном распределении;

в) параметра p в биномиальном распределении с n испытаниями. Являются ли полученные оценки несмещенными, состоятельными?

21. Для того чтобы узнать, сколько рыб в озере, отлавливают 500 рыб, метят их и выпускают обратно в озеро. Через некоторое время производится повторный отлов рыбы и среди 70 пойманных рыб оказываются 3 меченые рыбы. Оценить число рыб в озере.

22. Непрерывная случайная величина X распределена равномерно на отрезке $[a, a+1]$, где число a неизвестно. Для оценивания параметра a по простой выборке $X_1, X_2, ..., X_n$ генеральной совокупности X предлагаются две статистики:

$$A_1^* \text{ и } A_2^* : A_1^* = \overline{X} - 1/2;\ A_2^* = X_{(n)} - n/(n+1).$$

Являются ли они состоятельными и несмещенными? Какую из двух статистик использовать более целесообразно?

23. Являются ли достаточными следующие статистики: *а*) выборочное среднее $\overline{X}$ относительно параметра λ распределения Пуассона; *б*) частота "успехов" $n_{\text{усп}}/n$ относительно параметра p биноминального распределения; *в*) величина, обратная выборочному среднему: $1/\overline{X}$ относительно параметра μ показательного распределения; *г*) выборочное среднее $\overline{X}$ относительно параметра m нормального распределения (при известном параметре σ^2, при неизвестном параметре σ^2); *д*) выборочная дисперсия S^2 относительно параметра σ^2 нормального распределения (при известном параметре m, при неизвестном параметре m)?

24. Пусть $X_1, \ldots, X_n$ – простая выборка случайной величины X с равномерным распределением на отрезке $[0, \theta]$. Доказать, что порядковая статистика $X_{(n)}$ – полная достаточная статистика для θ и $T^* = \dfrac{n+1}{n} X_{(n)}$ – оптимальная несмещенная оценка θ.

25. Пусть $X_1, \ldots, X_n$ – простая выборка случайной величины X с равномерным распределением на отрезке $[\theta_1, \theta_2]$. Найти достаточную статистику:

а) относительно параметра θ_1,

б) относительно параметра θ_2,

в) относительно вектора $\boldsymbol{\theta} = (\theta_1, \theta_2)'$.

26. Пусть $X_1, \ldots, X_n$ – простая выборка случайной величины X с равномерным распределением на отрезке $[\theta_1, \theta_2]$. Доказать достаточность и полноту статистики $T = \left(X_{(1)}, X_{(n)}\right)'$ для векторного параметра $\theta = (\theta_1, \theta_2)'$. Найти оптимальные оценки для θ_1 и θ_2.

27. Пусть $X_1, \ldots, X_n$ – простая выборка случайной величины X, имеющей распределение Бернулли с параметром p. Доказать, что статистика $S = \sum_{i=1}^{n} X_i$ – полная достаточная статистика относительно p.

28. Испытывают n приборов. Считается, что время службы одного прибора до отказа – это экспоненциально распределенная случайная величина с параметром θ. Найти оценку максимального правдоподобия для параметра θ, если а) испытания проводят до отказа всех приборов, б) если испытания проводят до момента k-го отказа ($k < n$). Проверить достаточность и несмещенность полученных статистик (оценок).

29. Пусть $X_1, \ldots, X_n$ – простая выборка из равномерного на $[\theta, 2\theta]$ распределения. Найти достаточную статистику минимальной размерности.

30. Сталеплавильный завод изготовляет сталь, которая должна содержать 40% ванадия. Контроль содержания ванадия ведется на уровне значимости $\alpha = 0{,}05$. Методика контроля дает нормальное распределение результатов без систематической ошибки и со среднеквадратическим отклонением 2%. Контрольный

анализ конкретной партии стали дал для содержания ванадия 36,4%. Следует ли на основании полученного результата забраковать данную партию стали?

31. «Симметричная» монета бросается N раз. Найти функцию распределения числа X выпадений «герба» и оценить вероятность принадлежности частоты выпадения «герба» интервалу

$$(1/2 - \Delta,\ 1/2 + \Delta).$$

32. Измерительный прибор не имеет систематической ошибки, а случайная ошибка ζ имеет нормальное распределение: $\xi \in \mathsf{N}(0,\ \sigma^2)$, где величина σ^2 неизвестна. Оценить число n измерений $\{X_i\}_{i=1}^n$ случайной величины X, при котором оценка величины

$$\sigma^2 : S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

отличается от истинного значения σ^2 не более чем на 20% с вероятностью, не меньшей 0.7.

33. Известно, что событие A появляется в опыте с вероятностью $P(A) = 0{,}4$. Найти интервал минимальной длины, в котором лежит число наступлений события A в серии из восьми опытов: *а*) с вероятностью, не меньшей 0,6; *б*) с вероятностью, не меньшей 0,9.

34. Построить статистику для доверительного оценивания параметра λ в показательном распределении по простой выборке объема n.

35. Пятикратное измерение некоторой физической величины W одним и тем же прибором дало результаты: 1,78; 1,81; 1,94; 1,86; 2,00. Тем же прибором было произведено пятикратное измерение эталона, истинная величина которого равна одной единице измерения прибора. Результаты измерения эталона есть: 0,92; 0,78; 0,89; 0,82; 0,92. Предполагая, что ошибки измерений независимы и имеют одно и то же нормальное распределение, построить доверительный интервал для значений величины W при доверительной вероятности 0,95 (систематическая ошибка в обеих сериях измерений одинакова).

36. За первый час счетчиком зарегистрировано 150 событий пуассоновского потока, за следующие два часа – 250 событий. Была ли постоянной интенсивность наступления событий в единицу времени в течение всех трех часов наблюдения (уровень значимости α принять равным 0,05)?

37. По трем измерениям нормально распределенной случайной величины находится выборочное среднее $\overline{x} = 18{,}6$. Доверительная вероятность полагается равной 0,95. Найти доверительный интервал для значения математического ожидания при дисперсии, равной 0,25, считая ее: *а*) генеральной (истинной); *б*) выборочной (построенной на основании сделанных трех измерений). Сравнить результаты.

38. Построить оценку максимального правдоподобия параметра θ по простой выборке $X_1, \ldots, X_n$, где случайная величина X_i равномерно распределена на $[\theta, 2\theta]$.

39. Пусть о простой выборке $X_1, ..., X_n$ случайной величины X с распределением Пуассона ($X \in \mathsf{P}_0(\lambda)$) известно, что n_0 измерений равны нулю. Найти оценку максимального правдоподобия параметра λ.

40. В результате наблюдения точечного случайного процесса (потока событий) получена выборка ($X_1, \ldots, X_n$) моментов появления в нем событий. Предполагая, что наблюдаемый процесс является пуассоновским, найти МНП-оценки для интервала времени между событиями и для интенсивности потока событий.

41. По выборке объема n, извлеченной из нормальной двумерной генеральной совокупности (X, Y), где $MX = MY = 0$, $DX = DY = 1$, найден выборочный коэффициент корреляции $\hat{r}_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$. Требуется при уровне значимости α проверить гипотезу о некоррелированности случайных величин X и Y (т.е. равенстве нулю генерального (истинного) коэффициента корреляции r_{XY}). Рассмотреть случай

$$n = 1000,\ r_{\text{в}} = 0.3,\ \alpha = 0.01\,.$$

42. Функция $y = ax$ при неизвестном параметре a измерена в каждой из r точек x_i по n_i раз, $i = 1, 2, ..., r$. Пусть E_{ij} – случайная ошибка измерения, и результат измерений (x_i, y_{ij}) является реализацией уравнений $Y_{ij} = ax_i + E_{ij}$, $\left(i = \overline{1,r},\ \overline{j,n_i}\right)$. Полагая, что результаты измерений не коррелированы и $ME_{ij} = 0, DE_{ij} = \sigma^2$, найти оценку A^* параметра a, используя метод наименьших квадратов. Найти математическое ожидание и дисперсию оценки A^*.

43. В задаче № 42 обозначим $\overline{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$, $i = 1, 2, ..., r$. Подобрать постоянные $c_1, c_2, ..., c_n$ так, чтобы несмещенная оценка $A^{**} = \sum_{i=1}^{k} c_i \overline{y}_i$ параметра a имела наименьшую дисперсию. Найти дисперсию A^{**} при таком наилучшем выборе постоянных $c_1, c_2, ..., c_r$.

44. При изучении некоторого физического явления в термостате получены данные (в градусах Цельсия): 21,2; 21,8; 21,3; 21,0; 21,4; 21,3. Результаты измерений суть значения, принимаемые нормальными случайными величинами. К термостату применено некоторое усовершенствование, после чего на другом режиме получены данные (в градусах Цельсия): 37,7; 37,6; 37,6; 37,4. Можно ли при уровне значимости $\alpha = 0{,}05$ усовершенствование признать эффективным?

45. В течение всего апреля месяца сравнивались результаты работы предприятия в дневную и ночную смены. Получено, что в среднем за данный месяц в одну дневную смену производилось продукции на 62,7 тыс. руб., а в одну ночную смену – на 62,4 тыс. руб. Выборочные дисперсии. указанных объемов производства составили 0,66 для дневной смены и 0,80 для ночной смены. Предполагается, что объем производства за одну смену суть нормально распределенная случайная величина. Можно ли при уровне значимости $\alpha = 0{,}05$ считать, что объемы производства за одну смену в обеих сменах одинаковы?

46. Двумя методами проведены измерения одной и той же величины X и получены следующие результаты:

$x_{11} = 9,6;\ x_{12} = 10,0;\ x_{13} = 9,8;\ x_{14} = 10,2;\ x_{15} = 10,6$ (первый метод) и $x_{21} = 10,4;\ x_{22} = 9,7;\ x_{23} = 10,0;\ x_{24} = 10,3$ (второй метод).

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений (предполагается, что X суть нормальная случайна величина и полученные результаты суть реализации простых выборок)?

47. Результаты измерений значений нормально распределенной случайной величины Y при десяти значениях 1, 2, 3, ..., 10 неслучайной величины x есть 2,2; 3,1; 4,1; 5,0; 5,8; 6,9; 7.8; 9,0; 10,2; 11,1 соответственно. Построить уравнение линейной регрессии. Проверить гипотезу о целесообразности уточнения полученного уравнения в случае известной дисперсии случайной величины Y: $DY = 0,02$. Уточняет ли член $0,01\,x^2$ полученное уравнение регрессии?

48. Построить решающее правило, соответствующее критерию максимума апостериорной вероятности, для выбора по двум независимым измерениям одной из двух простых гипотез:

$H_1 : X \in \mathsf{N}(0,\ \sigma^2),\quad H_2 : X \in \mathsf{N}(1, 4\sigma^2)$

при $\sigma^2 = 1,\ P(H_1) = 0,4,\ P(H_2) = 0,6$. Выразить величины вероятностей ошибок первого и второго рода.

49. Построить решающее правило, соответствующее критерию максимума апостериорной вероятности, для выбора по двум независимым измерениям одной из двух простых гипотез:

$$H_1 : f\left(x \mid H_1\right) = (8\pi)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{8}\right),$$

$$H_2 : f\left(x \mid H_2\right) = 0,4(2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(x-1)^2}{2}\right) +$$

$$+ 0,6(2\pi)^{-\frac{1}{2}} \exp\left(-\frac{(x+1)^2}{2}\right)$$

при $P(H_1) = 0,3,\ P(H_2) = 0,7.$ Выразить величины вероятностей ошибок первого и второго рода.

50. До проведения эксперимента считалось, что случайная величина X может иметь одно из двух распределений: с вероятностью 0,2 – биномиальное с параметрами $n = 6,\ p = 0,2$; с вероятностью 0,8 – пуассоновское с параметром $\lambda = 3$. В результате четырех независимых измерений случайной величины X получены следующие результаты:

$$x_1 = 2,0;\ x_2 = 5,0;\ x_3 = 3,0;\ x_4 = 1,0.$$

Какое из распределений более вероятно?

51. До проведения эксперимента считалось, что случайная величина X может иметь одно из двух распределений: *а*) с вероятностью 0,4 – нормальное с параметрами $m = 2,\ \sigma^2 = 2$, *б*) с вероятностью 0,6 показательное с математическим ожиданием, равным 3. В результате четырех независимых измерений значений случайной величины X получены следующие результаты: $x_1 = 1,0;\ x_2 = 3,0;\ x_3 = 2,0;\ x_4 = 5,0.$ Какое из распределений более вероятно?

52. Построить байесовское решающее правило для выбора по двум независимым измерениям одной из двух простых гипотез:

$$H_1 : X \in f(x \mid H_1) = \mathrm{N}(-1, 1);$$

$$H_2 : X \in f(x \mid H_2) = \begin{cases} 2\exp(-2x), & \text{если } x \geq 0, \\ 0, & \text{если } x < 0 \end{cases}$$

при следующих условиях: $P(H_1) = 0,6;\ P(H_2) = 0,4;$ штраф за любое неправильное решение равен 2, штраф за верное решение равен –1. Найти величины вероятностей ошибок первого и второго рода.

53. Построить решающее правило, соответствующее критерию Неймана–Пирсона, для выбора по одному измерению при $\alpha = 0,1$ одной из двух гипотез

$$H_1:\ X \in \mathsf{N}(-1,1),\ \ H_2:\ X \in \mathsf{N}(2,4).$$

Определить величину β вероятности ошибки второго рода. Построить зависимость $\beta(\alpha)$.

54. Построить решающее правило, соответствующее критерию Неймана–Пирсона, для выбора по двум независимым измерениям при $\alpha = 0{,}02$ одной из двух гипотез:

$$H_1: X \in f\,(x \mid H_1) = \begin{cases} \dfrac{1}{3}, & \text{если } x \in [0,3], \\ 0, & \text{если } x \notin [0,3], \end{cases}$$

$$H_2: X \in f(x \mid H_2) = \begin{cases} 0{,}5\exp(-0{,}5x), & \text{если } x \geq 0, \\ 0, & \text{если } x < 0. \end{cases}$$

Найти величину вероятности ошибки второго рода.

55. Пусть $X_1,\ X_2,\ \ldots,\ X_n$ – выборка из биномиального распределения $\mathsf{Bi}(n,p)$. Построить критерий Неймана-Пирсона для проверки гипотезы $H_0: p = p_0$ против альтернативы $H_1: p = p_1 \left(0 < p_0 < p_1 < 1\right)$. и вычислить зависимость мощности критерия $\varphi = 1-\beta$ от допустимого значения вероятности ошибки первого рода α (рассмотреть нерандомизированное и рандомизированное решающие правила).

56. Для классификации состояния объекта, который может находиться в одном из двух состояний H_1 или H_2, может использоваться один из двух скалярных признаков Y_1 или Y_2, представляющих собой нормально распределенные случайные величины, связанные с состояниями объекта известными распределениями:

$$Y_1 \in \mathsf{N}(m_1(H_i), \sigma_1^2(H_i)),\ \ Y_2 \in \mathsf{N}(m_2(H_i), \sigma_2^2(H_i)),\ \ i = 1,2.$$

Для классификации состояния объекта используется критерий Неймана–Пирсона. Следует установить, какой из этих признаков предпочтительней, если параметры их распределений имеют следующие значения:

$$Y_1:\ m_1(H_1) = m_1(H_2) = 0,\ \ \sigma_1^2(H_1) = 1,\ \ \sigma_1^2(H_2) = 16;$$

$$Y_2:\ m_2(H_1) = 0,\ \ m_2(H_2) = 2,\ \ \sigma_2^2(H_1) = \sigma_2^2(H_2) = 1.$$

Для этого найдите для признаков с этими параметрами зависимости $\varphi = 1 - \beta = \varphi(\alpha)$. В частности, сравните эффективность признаков при допустимых значениях вероятности ошибки первого рода $\alpha_1 = 0{,}01$ и $\alpha_2 = 0{,}15$.

57. Случайные величины $X_1,\ X_2,\ \ldots,\ X_n$ независимы и одинаково распределены по показательному закону с неизвестным по величине параметром a. Построить критерий для проверки гипотезы $H_1 : a = a_0$ при альтернативной гипотезе $H_2 : a > a_0$, где a_0 – известное положительное число.

58. Построить критерий для проверки гипотезы $H_1 : p = \dfrac{1}{2}$ при альтернативной гипотезе $H_2 : p \neq \dfrac{1}{2}$ по результатам восьми испытаний, подчиняющихся схеме Бернулли. Вероятность ошибки первого рода α положить равной 0,05.

59. Пять независимых одинаково нормально распределенных случайных величин приняли значения: 3,02; 2,96; 3,06; 3,07; 2,96 соответственно. Проверить гипотезу $H_1 : \sigma^2 = 0{,}0036$ при альтернативе $H_2 : \sigma^2 > 0{,}0036$ при вероятности ошибки первого рода, равной 0,01.

60. Пять независимых одинаково распределенных случайных величин приняли значения: –0,46; +0,11; –0,32; +0,19; –0,17. Проверить гипотезу H_1:

$$H_1 : X \in f(x \mid H_1) = \begin{cases} 1, \text{ если } x \in [-0{,}5;\ 0{,}5], \\ 0, \text{ если } x \notin [-0{,}5;\ 0{,}5] \end{cases}$$

при альтернативе $H_2 : X \in \mathsf{N}(0,\ 0.03)$. Величину вероятности ошибки первого рода α положить равной 0,06. Найти вероятность ошибки второго рода β.

61. Найти статистику Z_n критерия Неймана–Пирсона для различения по простой выборке $X_1,\ X_2,\ \ldots,\ X_n$ гипотез:

$$H_0:\ X_k \in \mathrm{N}(\alpha_0,\ \sigma^2),\ k=1,\ 2,\ \ldots,\ n \quad \text{и}$$

$$H_1:\ X_k \in \mathrm{N}(\alpha_1,\ \sigma^2),\quad k=1, 2, \ldots, n,$$

где $\alpha_0,\ \alpha_1, \sigma^2$ – заданные числа.

62. Найти статистику Z_n критерия Неймана–Пирсона для различения по простой выборке $X_1, X_2, \ldots, X_n$ гипотез:

$$H_0: P\{X_k = t\} = p_t^{(0)},\quad t=1,\ 2,\ \ldots,\ N;\quad k=1,\ 2,\ \ldots,\ n \ \text{и}$$

$$H_1: P\{X_k = t\} = p_t^{(1)},\quad t=1,\ 2,\ \ldots,\ N;\quad k=1,\ 2,\ \ldots,\ n,$$

где $p_1^{(0)}, \ldots, p_N^{(0)},\quad p_1^{(1)}, \ldots, p_N^{(1)}$ – заданные числа,

$$\sum_{t=1}^{N} p_t^{(0)} = \sum_{t=1}^{N} p_t^{(1)} = 1.$$

63. Пусть $X \in \mathrm{N}(m, 1)$, где $P\{m=1\} = p$ или $P\{m=-1\} = 1-p$. Построить байесовское решающее правило для различения двух возможных значений m при единичном (нулевом) штрафе за ошибочное (правильное) решение.

64. Значение сигнала Y на входе некоторого устройства может быть либо нулем, либо единицей. Значение Y недоступно для измерения. На выходе устройства наблюдается (и измеряется) величина X, являющаяся суммой входного сигнала и гауссовского шума с нулевым математическим ожиданием и известной дисперсией σ^2. Построить оптимальное байесовское решающее правило для классификации входных сигналов на основании измерения величины X при известных вероятностях

$$P\{Y=0\} = p,\ P\{Y=1\} = 1-p.$$

65. Пусть гипотезы H_1 и H_2 относительно распределения двухмерного случайного вектора $\boldsymbol{X} = (X_1, X_2)'$ имеют вид

$$H_1: X \in \mathrm{N}(m_1, R_1),\quad H_2: X \in \mathrm{N}(m_2, R_2),$$

где $$m_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\quad m_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix},\quad R_1 = R_2 = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix},$$

$$P(H_1) = P(H_2) = \frac{1}{2}.$$

Найти и изобразить графически: *а*) разделяющую границу, минимизирующую величину вероятности $P_{\text{ош}}$ ошибочного решения по различению гипотез H_1 и H_2; *б*) разделяющую границу, минимизирующую средний риск при следующих значениях штрафов: $c_{11} = c_{22} = 0$, $c_{12} = 2\,c_{21}$. Оценить зависимость $P_{\text{ош}}$ от ошибочного неучета коррелированности компонент X_1 и X_2 вектора $\boldsymbol{X}$.

66. Рассматриваются те же гипотезы H_1 и H_2, что и в задаче № 65, при $c_{11} = c_{22} = 0$, $c_{21} = c_{12}$, $P(H_1) = P(H_2)$. Найти минимальные значения среднего риска, если решение принимается на основании измерения

а) только первой компоненты X_1 вектора $\boldsymbol{X}$;

б) только второй компоненты X_2 вектора $\boldsymbol{X}$;

в) вектора $\boldsymbol{X}$.

67. Пусть гипотезы H_1 и H_2 относительно распределения двухмерного случайного вектора $\boldsymbol{X} = (X_1, X_2)'$ имеют вид

$$H_1 : X \in N(m_1, R_1), \quad H_2 : X \in N(m_2, R_2),$$

где $$m_1 = m_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad R_1 = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}.$$

Изобразить графически разделяющую границу, минимизирующую средний риск при следующих значениях штрафов: $c_{11} = c_{22} = 0$, $c_{12} = c_{21} = 1$. Рассмотреть случаи:

$$P(H_1) = 0.5P(H_2), \; P(H_1) = P(H_2) = 0.5.$$

Исследовать зависимость классифицирующей способности вектора $\boldsymbol{X}$ от значения коэффициента корреляции его компонент. Какова индивидуальная классифицирующая способность его компонент?

68. Необходимо произвести выбор между двумя гипотезами о возможных значениях p_0 и p_1 вероятности события

A $(p_0 < p_1)$. В этих целях осуществляется последовательность независимых опытов, в каждом из которых определяется, происходит или не происходит событие A. Построить последовательный критерий отношения вероятностей при заданных значениях α и β вероятностей ошибок первого и второго рода.

69. Пусть X – число успехов в n независимых испытаниях с вероятностью успеха p (неизвестной) в каждом испытании. Построить критерий, имеющий мощность не менее 0,9, для проверки гипотезы $H_0 : p = 0,2$ при альтернативе $H_1 : p = 0,4$ для заданной величины вероятности ошибки первого рода $\alpha = 0,05$. Определить необходимый объем выборки: *а*) используя таблицы биноминального распределения; *б*) используя нормальное приближение.

70. Пусть гипотезы H_0 и H_1 имеют вид

$$H_0 : f(x \mid H_0) = \theta_0^{-1} \exp(-x / \theta_0), \quad x > 0;$$

$$H_1 : f(x \mid H_1) = \theta_1^{-1} \exp(-x / \theta_1), \quad x > 0; \quad \theta_0 = 2\theta_1.$$

Построить процедуру различения гипотез H_1 и H_2 с фиксированным объемом выборки при заданных величинах вероятностей ошибок первого и второго рода $\alpha = \beta \le 0,05$ и процедуру последовательного критерия отношения вероятностей при тех же значениях α и β. Сравнить необходимые в обоих случаях объемы выборок.

71. Используя критерий Неймана–Пирсона для одноэлементной выборки, построить оптимальное решающее правило для проверки гипотезы H_1 о наблюдении случайной величины X_1 с распределением вероятностей:

$$P(X_1 = -1) = 0,7, \quad P(X_1 = 0) = 0,25, \quad P(X_1 = 1) = 0,05$$

при альтернативной гипотезе H_2 о наблюдении случайной величины $X_2 = -X_1$ при ограничении на вероятность ошибки первого рода $\alpha \le \alpha^* = 0,01$. Применить и сравнить нерандомизированное и рандомизированное решающие правила. Рассмотреть зависимость $\beta = \beta(\alpha)$ для таких правил при $\alpha^* = \text{var}$.

72. Имеется простая выборка $X_1, X_2, \ldots, X_n$. По гипотезе H_0 все элементы выборки равномерно распределены на отрезке [0, 2], а по гипотезе H_1 – на отрезке [1, 3]. Построить критерий для различия гипотез H_0, H_1 с наименьшей величиной $\max(\alpha, \beta)$, где α и β – вероятности ошибок первого и второго рода соответственно.

73. Решить задачу № 72, если по гипотезе H_1 все элементы выборки равномерно распределены на отрезке [1, 4].

Дополнительные задачи
по курсу «Математическая статистика»

1. Предположим, что независимо n раз кидается монетка с вероятностью выпадения орла в каждом опыте, равной p, т.е. Сколько раз нужно кинуть монетку (оцените n), чтобы оценка (выборочное среднее) $\overline{p}(\vec{x}) = \frac{1}{n}\sum_{k=1}^{n} x_k$, где $x_k \in Be(p)$, с вероятностью, не меньшей, чем $\gamma = 0,95$, отличалась от истинного значения p не более чем на величину $\delta = 0,01$? Применить неравенство Чебышева и предельную теорему (точность, которую дает ц.п.т., оцените с помощью неравенства Берри–Эссена). Сравнить результаты. Является ли $\overline{p}(\vec{x})$ оценкой максимального правдоподобия? Состоятельной? Оптимальной (эффективной)?

2. В некоторой стране прошел второй тур выборов. Выбор был между двумя кандидатами А и В (графы “против всех” на этих выборах не было). Сколько человек надо опросить на выходе с избирательных участков, чтобы, исходя из ответов, можно было определить долю проголосовавших за кандидата А с точностью 1% и с вероятностью, не меньшей 0,9?

3. Любопытный студент швейного техникума решил повторить опыты Бюффона по бросанию иглы (студент хочет оценить число π). Для этого он подготовил горизонтально расположенный лист бумаги, разлинованный параллельными прямыми так, что расстояние между соседними прямыми равно 1. Однако в распоряжении студента оказалась только погнутая иголка. Иголка имеет форму кочерги, но студент не имеет точного представления о том, как именно погнута иголка. Ему известно лишь то, что длина иголки равна 2. Студент бросил погнутую иголку 1 000 000 раз и посчитал суммарное число пересечений, учитывая кратность. Помогите студенту оценить число π: а) с помощью неравенства Чебышева; б)* с помощью з.б.ч. (закона больших чисел) и неравенств о вероятностях больших уклонений; в) с помощью ц.п.т. (центральной предельной теоремы) и оценок скорости сходимости в ц.п.т., например, с помощью неравенства Берри–Эссена или более точных аппроксимаций.

4. В некотором вузе проходит экзамен. Количество экзаменационных билетов $N \gg 1$. Перед экзаменационной аудиторией выстроилась очередь из студентов, которые не знают чему равно N. Согласно этой очереди студенты вызываются на экзамен (второй студент заходит в аудиторию, после того как из нее выйдет первый, и т.д.). Каждый студент с равной вероятностью может выбрать любой из N билетов (в независимости от других студентов). Проэкзаменованные студенты, выходя из аудитории, сообщают оставшейся очереди номера своих билетов. Оцените (сверху), сколько студентов должно быть проэкзаменовано, чтобы оставшаяся к этому моменту очередь смогла оценить число экзаменационных билетов с точностью 10% с вероятностью, не меньшей 0,95.

5. (Метод Монте-Карло.) Используя методы математической статистики, предложите эффективный способ вычисления с заданной точностью ε и с заданной доверительной вероятностью γ абсолютно сходящегося интеграла $J = \int\limits_{[0,1]^m} f(\vec{x})d\vec{x}$. Считайте, что $\forall\ \vec{x} \in [0,1]^m \to |f(\vec{x})| \le 1$.

Пояснение. Введем случайный m-вектор $\vec{X} \in R\left([0,1]^m\right)$ и с.в. $\xi = f(\vec{X})$. Тогда $M\xi = \int\limits_{[0,1]^m} f(\vec{x})d\vec{x} = J$. Поэтому получаем оценку интеграла $\overline{J}_n = \frac{1}{n}\sum\limits_{k=1}^{n} f(\vec{x}^k)$, где $\vec{x}^k$, $k = 1,...,n$ – повторная выборка значений случайного вектора $\vec{X}$ (т.е. все $\vec{x}^k$, $k = 1,...,n$ – независимы и одинаково распределены: так же как и вектор $\vec{X}$). В задаче требуется оценить сверху число n ($n \gg m$), начиная с которого $P\left(|J - \overline{J}_n| \le \varepsilon\right) \ge \gamma$.

6. (Распределение Коши.) На плоскости на расстоянии $a > 0$ (неизвестный параметр) от детектирующей прямой располагается радиоактивный источник, который излучает вспышками равновероятно по любому направлению в этой плоскости. Пусть $\vec{x} = (x_1,...,x_n)$ – вектор координат вспышек, регистрируемых де-

тектором. Требуется построить по такой простой выборке состоятельную оценку координаты проекции источника на детектирующую прямую.

7. Пусть имеется простая выборка $\vec{x} = (x_1,...,x_n)$ из равномерного распределения на отрезке $[0,1]$, т.е. $x_k \in R[0,1]$. Как из этой выборки получить простую выборку (того же объема) из стандартного нормального распределения $N(0,1)$? Как из простой выборки из распределения $Be(1/2)$ получить простую выборку из распределения $R[0,1]$? Считайте длину мантиссы μ конечной, т.е. любое действительное число из отрезка $[0,1]$ мы округляем, оставляя в двоичным представлении этого числа лишь фиксированное число слагаемых μ.

8. Пусть имеется простая выборка $\vec{x} = (x_1,...,x_n)$ из распределения с неизвестной гладкой плотностью $p(x) > 0$, сосредоточенной на некотором отрезке. Покажите, что для оценки плотности распределения методом гистограмм с одинаковой длиной интервалов Δ (где функция плотности распределения на рассматриваемом интервале приближается константой, равной числу наблюдений, попавших в этот интервал, нормированной на общее число наблюдений и ширину интервала) “оптимально” выбирать число интервалов $n_{opt} \sim n^{1/3}$ (оценка Н. В. Смирнова). При этом относительная погрешность такой аппроксимации[1] оценивается как

[1] Квадрат суммарной относительной погрешности в данном интервале δ^2 складывается (дисперсия суммы равна сумме дисперсий) из квадрата относительной статистической погрешности ${\delta_s}^2$ и квадрата относительной погрешности аппроксимации непрерывно меняющейся функции константой (для метода гистограмм) или линейной функцией (для полигона частот) ${\delta_a}^2$. Для определения ${\delta_s}^2$ можно воспользоваться пуассоновским приближением: ${\delta_s}^2 = \left(np(x)\Delta\right)^{-1}$.

$\sqrt{\delta^2_{\min}} \sim n^{-1/3}$. Если вместо метода гистограмм использовать метод полигона частот (при котором прямыми линиями соединяются середины прямоугольных ступенек и затем неизвестная плотность приближается получающимися трапециями), то $\sqrt{\delta^2_{\min}} \sim n^{-2/5}$. Отметим, что такую же оценку дает и ядерный метод оценки плотности Розенблатта–Парзена.

9. В методе Н. Н. Ченцова неизвестная плотность $p(x)$ ищется в виде ряда $p(x) = \sum\limits_{k=1}^{\infty} c_k \varphi_k(x)$ по ортонормированной в $L_2(\mathbb{R}, p(x))$ системе функций $\{\varphi_k(x)\}_{k=1}^{\infty}$. При этом оценки c_k вычисляются по простой выборке $\vec{x} = (x_1, ..., x_n)$ из распределения с плотностью $p(x)$ с помощью формулы $\hat{c}_k = n^{-1} \sum\limits_{i=1}^{n} \varphi_k(x_i)$. Покажите, что оценка относительной погрешности по методу Ченцова (с m базисными функциями) имеет вид $\sqrt{\delta^2_{\min}} \sim mn^{-1/2}$. Заметим, что не существует методов оценки неизвестной функции плотности, имеющих большую скорость убывания погрешности по n.

10. Пусть $\vec{X} = (X_1, X_2, ..., X_n)$ – простая выборка объема n из распределения $F(x)$. Покажите, что в случае, когда $F(x)$ – непрерывная функция, распределение статистики

$$D_n(\vec{X}) = \sup_{x \in \mathbb{R}} \left| F_n(x; \vec{X}) - F(x) \right|$$

не зависит от того, какая именно функция $F(x)$.

Здесь $F_n(x; \vec{x})$ – эмпирическая функция распределения:

$$F_n(x; \vec{X}) = \frac{\mu_n(x)}{n} = \frac{1}{n} \sum_{k=1}^{n} I_k(x),$$

$$\mu_n(x) = \left| \{ j : X_j < x \} \right|, \; I_k(x) = \begin{cases} 1, & X_k < x, \\ 0, & X_k \geq x. \end{cases}$$

В действительности, подобный результат будет справедлив и для широкого класса статистик $G(F_n, F)$ (т.е. измеримого функционала) от $F_n\left(x; \vec{X}\right)$.[2]

---

[2] Если говорить точнее, то имеет место следующий результат (см. книгу [4]). Пусть имеется выборка $\vec{x}$ из распределения с непрерывной функцией распределения $F_0(x)$. И задан функционал $G(F)$ такой, что

$$\forall\ f \in C(\mathbb{R}),\ \{f_h\}_h \in D(0,1):\ f_h \xrightarrow[h\to 0]{} f\ \ \exists\ G'(F_0, f):$$

$$\frac{G(F_0 + hf_h) - G(F_0 + hf_h)}{h^k} \xrightarrow[h\to 0]{} G'(F_0, f),$$

$$G'(F_0, f_h) \xrightarrow[h\to 0]{} G'(F_0, f),$$

где $D(0,1)$ – пространство функций на $[0,1]$, непрерывных слева (и справа в точке 1) и имеющих лишь конечное число скачков. Тогда

$$n^{k/2}\left(G\left(F_n(x;\vec{x})\right) - G(F_0)\right) \xrightarrow[n\to\infty]{d} G'\left(F_0, w^0(F_0)\right),$$

где d означает сходимость по распределению, а броуновский мост $w^0(t)$ $t \in [0,1]$ определяется как $w^0(t) = W(t) - tW(1)$, где $W(t)-$ винеровский процесс. Отсюда, например, легко получается основная теорема критерия согласия Колмогорова и теорема Пирсона (о сходимости к χ^2).

Если относительно $G(F)$ известно, что $G(F) = Q\left(\int g(x)dF(x)\right)$ (можно обобщить и на случай нескольких аргументов вида $\int g_l(x)dF(x)$), где $\int g^2(x)dF(x) < \infty$ и $G(F)$ непрерывно дифференцируем по Фреше в $a_0 = \int g(x)dF_0(x)$, то

$$\sqrt{n}\left(Q\left(\frac{1}{n}\sum_{k=1}^{n} g(x_k)\right) - Q(a_0)\right) \xrightarrow[n\to\infty]{d}$$

$$\xrightarrow[n\to\infty]{d} Q'(a_0)\cdot N\left(0, \int\left(g(x) - a_0\right)^2 dF_0(x)\right).$$

Указание. Положим по определению

$$F^{-1}(u) = \inf\{\xi : F(\xi) = u\} = \min\{\xi : F(\xi) = u\} \text{ для } u \in [0,1],$$

где последнее равенство имеет место в силу непрерывности $F(x)$. Понятно, что это отнюдь не единственный способ выбора однозначной функции из, вообще говоря, многозначного отображения $F^{-1}(u)$, однако именно такое определение окажется наиболее полезным в дальнейшем. Положим $u = F(x)$, тогда u пробегает как минимум все точки интервала $(0, 1)$ (а как максимум – отрезка $[0,1]$), когда x пробегает $\mathbb{R}$.[3] Делая замену $u = F(x)$ и используя определение $F^{-1}(u)$, получим $D_n(\vec{X}) = \max\limits_{u\in[0,1]} \left| F_n\left(F^{-1}(u); \vec{X}\right) - u \right|$.[4] Далее имеем

$$F_n\left(F^{-1}(u)\right) = \frac{1}{n}\sum_{k=1}^{n} \theta\left(F^{-1}(u) - X_k\right).$$

Остается показать эквивалентность событий

$$\left\{F^{-1}(u) - X_k \le 0\right\} \sim \left\{u - F(X_k) \le 0\right\}.$$

Сделайте это, завершив тем самым доказательство сформулированного утверждения.

11. Дана простая выборка $\vec{x} = (x_1, ..., x_n)$. Имеются две гипотезы:

$$H_0: \quad x_k \in R[0, \theta_0], \; H_1: \quad x_k \in R[0, \theta_1].$$

Постройте с уровнем значимости α наиболее мощный критерий (Неймана–Пирсона) проверки гипотезы H_0 против альтернативы H_1 в случае:

а) $\theta_0 < \theta_1$; б) $\theta_0 > \theta_1$.

---

[3] Это следует из непрерывности $F(x)$.

[4] Обратим внимание, что sup по интервалу мы заменили на max по отрезку, равна замыканию интервала.

12. *Экспериментатор располагает монеткой, относительно которой имеются две гипотезы: $H_0 : p = \frac{1}{3}$ и $H_1 : p = \frac{2}{3}$. Задавшись уровнями ошибок первого и второго рода $\alpha = \beta = 0.05$, предложите алгоритм, минимизирующий одновременно

$$M\left(N_{\alpha,\beta}\middle|H_0\right) \text{ и } M\left(N_{\alpha,\beta}\middle|H_1\right),$$

где с.в. $N_{\alpha,\beta}$ – число бросаний монетки до момента остановки. В момент остановки экспериментатор выбирает одну из альтернатив, при этом допускаемые им ошибки:

$$P\left(H_1\middle|H_0\right) \le \alpha,\ P\left(H_0\middle|H_1\right) \le \beta.$$

13. Даны k простых выборок (взаимно независимых) объемами n_1, …, n_k из распределений $Be(\theta_1)$, …, $Be(\theta_k)$. С помощью асимптотического $(n_1 \to \infty, \ldots, n_k \to \infty)$ критерия отношения правдоподобия (КОП) с уровнем значимости α проверить гипотезу однородности $H_0 :\ \theta_1 = \ldots = \theta_k$ (H_1 – H_0 не верна).

14. До проведения схемы испытаний Бернулли разыгрывается с.в. p, имеющая равномерное распределение на отрезке $[0{,}1;\ 0{,}9]$ (результаты розыгрыша нам неизвестны). После того как эта с.в. была разыграна, начинают проводиться опыты по схеме Бернулли (независимо $n = 1000$ раз подкидывается монетка) с вероятность успеха (выпадения «орла») в каждом опыте, равной p (после того как с.в. p была разыграна, она уже приняла какое-то значения из отрезка $[0{,}1;\ 0{,}9]$ и рассматривается в серии опытов Бернулли уже как число, причем не меняющееся от опыта к опыту). В результате опыта было посчитано значение числа успехов $r = 777$. Оцените по методу максимума апостериорной вероятности значение p. Как изменится ответ, если точное значение числа успехов нам неизвестно? Известно только, что $r \in [750, 790]$.

15. (Робастные оценки, *M*-оценки.)* Дана простая выборка $\vec{x} = (x_1, \ldots, x_n)$. Известно, что $x_k \in (1-\varepsilon)N(\theta,1) + \varepsilon N\left(0,\sigma^2\right)$ – считаем нормальные случайные величины в этой сумме независи-

мыми. Значение параметра $\sigma^2 > 0$ известно и считается довольно большим. Значение параметра $\varepsilon \geq 0$, отвечающего за "помехи", неизвестно, но считается, что $\varepsilon \leq \varepsilon_0$, где ε_0 – известно и достаточно мало. Постройте минимаксную оценку (П. Хьюбера) неизвестного параметра θ, о котором априорно ничего неизвестно. Если дополнительно искать оптимальную оценку в классе усеченных оценок или в классе *M*-оценок Винзора:

$$\hat{\theta}_1(\vec{x}) = \frac{1}{n-2m}\sum_{k=m+1}^{n-m} x_{(k)},$$

$$\hat{\theta}_2(\vec{x}) = \frac{1}{n}\left[\sum_{k=m+1}^{n-m} x_{(k)} + m\cdot\left(x_{(m+1)} + x_{(n-m)}\right)\right],$$

то разумно ли выбирать $m(\varepsilon) = [\varepsilon_0 n/2]$ для усеченной оценки? Как следует выбрать $m(\varepsilon_0)$ для оценки Винзора?

16. Пусть дано геометрическое распределение с параметром $\lambda \in [0{,}01;\ 0{,}99]$. В *n*-м опыте, где n пробегает натуральный ряд, экспериментатор может лишь ответить на вопрос: выполняется неравенство $x_n \leq m_n$ или нет? Исследовать с помощью МНП теоремы, если возможно, асимптотическое поведение МНП оценки, в случае когда

1) $m_n = \lambda$; 2)* $m_n = \left\lceil \left(\ln(2+n)\right) \big/ \left(\ln(1/\lambda)\right) \right\rceil$ и 3) $m_n = n$.

17. Пусть $\xi \in Po(\lambda), \lambda \gg 1$. Покажите, что $\sqrt{\xi} \overset{d}{\approx} \sqrt{\lambda} + N(0, 1/4)$. Исходя из этого свойства, постройте доверительный интервал для простой выборки $\vec{x} = (x_1, ..., x_n)$ из распределения $Po(\mu)$ с неизвестным параметром $\mu > 0$.

18. В модели Блэка–Шоулса–Мертона эволюция цены акции описывается геометрическим броуновским движением:

$$S(t) = S(0)\exp\left(at + \sigma W(t)\right),$$

где $W(t)$ – винеровский процесс $(\sigma > 0)$. С помощью эргодической теоремы для случайных процессов оцените неизвестный параметр a, если известна реализация процесса $S(t)$ на достаточно

длинном временном отрезке $[0, T]$. Предложите способ оценки неизвестного параметра σ.

19. Закон Хаббла в астрономии гласит: "скорость удаления галактики V прямо пропорциональна (с коэффициентом пропорциональности H – постоянная Хаббла) расстоянию до неё R". Будем считать, что ошибки измерений некоррелированы, не имеют систематической ошибки и одинаково распределены по нормальному закону, т.е. имеет место "нормальная регрессия":

$$\vec{V} = H\vec{R} + \vec{\varepsilon},\ \vec{\varepsilon} \in N\left(0, \sigma^2 I_n\right)\ (\vec{V} \text{ и } \vec{R} \text{ – известны}).$$

а) Предложите формулу для параметра H (ответ аргументируйте).
б) Постройте γ-доверительный интервал для параметра H, если σ^2 известно.
в) Постройте γ-доверительный интервал для параметра H, если σ^2 неизвестно.

20. (Метод спейсингов.) Пусть $\vec{X} = (X_1, X_2, ..., X_n)$ – простая выборка объема n из распределения $F(x, \theta)$, где θ – неизвестный параметр (возможно векторный). Один из способов оценивания θ заключается в следующем:

$$\hat{\theta} = \arg\max_{\theta} H(\theta),$$

где

$$H(\theta) = \sum_{i=1}^{n+1} \ln D_i(\theta) = \sum_{i=1}^{n+1} \ln\left(F\left(X_{(i)}, \theta\right) - F\left(X_{(i-1)}, \theta\right)\right).$$

С.в. $D_i(\theta),\ i = 1, \ldots, n+1,$ называют долями выборки $\vec{X}$.

Для выборки из равномерного распределения на $[\mu_1, \mu_2]$ получите оценки параметров μ_1 и μ_2:
а) методом максимального правдоподобия;
б) методом спейсингов.
Являются ли полученные оценки несмещенными?

21. Пусть $\vec{X} = (X_1, X_2, ..., X_n)$ – простая выборка из равномерного распределения на $[0, \theta]$. Используя точечную оценку $\hat{\theta} = X_{(n)}$

для параметра θ, постройте γ-доверительный интервал. Сравните с доверительным интервалом, построенным на основе центральной статистики.

22. (Гипотеза независимости двух выборок; метод ранговых сумм Уилкоксона.) Пусть $\vec{X} = (X_1, X_2, ..., X_n)$ – простая выборка из непрерывного распределения $F(x)$, $\vec{Y} = (Y_1, Y_2, ..., Y_m)$ – простая выборка из непрерывного распределения $G(x)$. Все компоненты случайного вектора

$$(X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_m)$$

независимы в совокупности. Для тестирования основной гипотезы

$$H_0 : \ \forall \ x \in \mathbb{R} \to F(x) = G(x)$$

против альтернативы доминирования

$$H_1 : \ \forall \ x \in \mathbb{R} \to F(x) \geq G(x)$$

(т.е. с.в. Y стохастически больше с.в. X)
воспользуйтесь асимптотически нормальной (при справедливости основной гипотезы) статистикой

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{\{X_i < Y_j\}}.$$

Указание.
Покажите, что если справедлива основная гипотеза, то

$$MU = \frac{nm}{2}, \ DU = \frac{nm(n+m+1)}{12}.$$

23. (Парадокс критерия хи-квадрат.) Ниже приведены три таблицы, в которых отражено действие некоторого лекарства (способа лечения) только на мужчин, только на женщин, и, наконец, на больных обоего пола (объединенные результаты).

Мужчины	Выздор.	Не выздор.
Приним. лек-во	700	800
Не приним. лек-во	80	130

Женщины	Выздор.	Не выздор.
Приним. лек-во	150	70
Не приним. лек-во	400	280

Вместе	Выздоровление	Не выздоровление
Принимали лекарство	850	870
Не принимали лекарство	480	410

Примените критерий хи-квадрат для тестирования гипотезы однородности (проверка эффективности лекарства) для каждой из таблиц.

Заметьте, что, судя по третьей таблице, доля выздоровевших среди тех людей, что не принимали лекарство, больше.

Объясните полученные результаты.

24. Карта южной части Лондона была разбита на $n = 24 \times 24 = 576$ небольших участков, каждый площадью 0,25 кв.км. На карте были отмечены места падения самолетов-снарядов во время Второй мировой войны. В таблице приведены количества участков l_j ровно с j падениями, $j = 0,1,...,7$.

j	0	1	2	3	4	5	6	7
l_j	229	211	93	35	7	0	0	1

Проверьте гипотезу о низкой точности стрельбы.

Комментарии. Имеется в виду, что в силу большого количества участков вероятность попадания на отдельный участок самолета-снаряда мала, значит, при справедливости гипотезы о низкой точности стрельбы можно воспользоваться законом редких событий, согласно которому число попаданий на любой из участков есть (приближенно) пуассоновская с.в. с некоторым общим для всех участков параметром θ. Попадания на разные участки независимы.

Указание. Прежде чем применять критерий хи-квадрат для сложной гипотезы, постарайтесь правильно разбить данные на интервалы (вспомните условия применения критерия), в качестве первоначальной оценки вероятности попадания в отдельный интервал воспользуйтесь приближением для неизвестного параметра

$$\tilde{\theta} = \frac{1}{n} \sum_{j=0}^{7} j l_j \approx 0.932.$$

25. (Метод «складного ножа» первого порядка.) Пусть $T_n\left(\vec{X}\right)$ – некоторая *смещенная* оценка параметра θ по выборке $\vec{X}=\left(X_1,X_2,...,X_n\right)$ объема n. Пусть смещение этой оценки имеет вид

$$MT_n\left(\vec{X}\right)-\theta=\frac{a_1}{n}+\frac{a_2}{n^2}+\ldots$$

Рассмотрим оценку

$$\hat{\theta}=nT_n\left(\vec{X}\right)-\frac{n-1}{n}\sum_{i=1}^{n}T_{n-1}^{(-i)}\left(\vec{X}\right),$$

где $T_{n-1}^{(-i)}\left(\vec{X}\right)=T_{n-1}\left(\vec{X}^{(-i)}\right)=T_{n-1}\left(X_1,...,X_{i-1},X_{i+1},...,X_n\right)$.

Покажите, что смещение этой оценки не содержит членов порядка $O\left(\frac{1}{n}\right)$.

Примените метод «складного ножа» для построения новой (вообще говоря, уже несмещенной) оценки дисперсии $\hat{\theta}_2{}^2$ по оценке максимального правдоподобия для выборки из нормального распределения $N\left(\theta_1,\theta_2{}^2\right)$.

26. Пусть с.в. $X_{ij}=\mu_j+\varepsilon_{ij}$, где $\varepsilon_{ij}\in N\left(0,\sigma^2\right)$, $i=1,...,n_j$, $j=1,...,k$ – независимые с.в. (k независимых простых выборок объема $n_1,...,n_k$), $N=\sum_{j=1}^{k}n_j$. Для проверки гипотезы $H_0:\ \mu_1=...=\mu_k$ (H_1 – H_0 не верна) с уровнем значимости α воспользуйтесь статистикой:

$$R=\frac{\frac{1}{k-1}\sum_{j=1}^{k}n_j\left(X_{\bullet j}-X_{\bullet\bullet}\right)^2}{\frac{1}{N-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij}-X_{\bullet j}\right)^2},$$

где $$X_{\bullet j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}, \quad X_{\bullet\bullet} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{ij}.$$

Покажите, что при справедливости гипотезы H_0 статистика имеет распределение $F_{k-1,N-k}$ (Фишера–Снедекора).

27. Пусть есть группа из n человек ($n \gg 1$), каждый из них может быть потенциально болен. Для выявления болезни человеку делают анализ крови.

Методика: смешиваются пробы крови k человек, и анализируется полученная смесь. Если антител нет, то одной проверки достаточно для k человек. В противном случае кровь каждого человека нужно исследовать отдельно, и для k человек всего потребуется $k+1$ раз провести анализ.

Вероятностная модель: предположим, что вероятность обнаружения антител p ($p \ll 1$) одна и та же для всех n обследуемых, и результаты анализов для различных людей независимы, т.е. моделью является последовательность из n испытаний Бернулли с вероятностью «успеха» p.

Покажите, что предложенная методика позволяет выявить всех больных при числе проверок (анализов) в среднем в несколько раз меньшем, чем общее число людей.

Указание. Определите размер группы $k_0 = k_0(p)$, минимизирующий среднее число проверок. Покажите, что если $p = 0{,}01$, то в среднем потребуется приблизительно в пять раз меньше проверок, чем общее число людей.

Теоретические вопросы

1. Получите обоснование критериев независимости и однородности с помощью доказательства теоремы Крамера (параметрический χ^2), приведенного в [6, п. 30.3].

2. Покажите, что в наиболее мощном решающем правиле (Неймана–Пирсона):

$$\varphi(\vec{x}) = \begin{cases} 1, & \Lambda(\vec{x}) > \overline{\Lambda} \\ p(\vec{x}), & \Lambda(\vec{x}) = \overline{\Lambda} \\ 0, & \Lambda(\vec{x}) < \overline{\Lambda} \end{cases}, \text{ где } \Lambda(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)},$$

$\overline{\Lambda}$ и $0 \le p(\vec{x}) \le 1$ следует определять из условия (ошибкой первого рода):

$$P(H_1|H_0) = \int\limits_{\{\vec{x}:\ \Lambda(\vec{x}) > \overline{\Lambda}\}} L(\vec{x}|H_0)\,d\vec{x} + \\ + \int\limits_{\{\vec{x}:\ \Lambda(\vec{x}) = \overline{\Lambda}\}} p(\vec{x})\,L(\vec{x}|H_0)\,d\vec{x} = \alpha. \qquad (*)$$

Причем $\overline{\Lambda}$ определяется единственным образом, а от того как выбирать $p(\vec{x})$, удовлетворяющее (*), не зависит ошибка второго рода:

$$P(H_0|H_1) = \int\limits_{\{\vec{x}:\ \Lambda(\vec{x}) < \overline{\Lambda}\}} L(\vec{x}|H_1)\,d\vec{x} + \\ + \int\limits_{\{\vec{x}:\ \Lambda(\vec{x}) = \overline{\Lambda}\}} \bigl(1 - p(\vec{x})\bigr)\,L(\vec{x}|H_1)\,d\vec{x} = \beta.$$

Всегда ли можно искать $p(\vec{x})$ в виде константы $p(\vec{x}) \equiv p$?

3. Поясните, почему в последовательном анализе Вальда "считают", что имеют место следующие приближенные равенства [1]:

$$\alpha \approx \frac{1-\beta}{\Lambda_2},\ 1-\alpha \approx \frac{\beta}{\Lambda_1}.$$

4. Формализуйте и обоснуйте следующее утверждение: "равномерно наиболее мощному критерию проверки простой гипо-

тезы против двусторонней сложной альтернативы соответствует равномерно наикратчайший доверительный интервал, и наоборот”.[5]

5. Всегда ли существует несмещенная оценка, МНП-оценка, эффективная оценка, оптимальная оценка? Приведите примеры.

6. Покажите, что если существует эффективная оценка, то МНП-оценка совпадает с ней.

7. а) Предложите способ построения асимптотически наикратчайших доверительных интервалов (областей) с помощью теоремы об асимптотических свойствах МНП оценок. б)* Как можно использовать теорему об асимптотических свойствах критерия отношения правдоподобия (КОП) из [2] для построения доверительных областей?

Исследовательские задачи

1. (Асимптотически оптимальные адаптивные правила.)[6] Имеется два “одноруких бандита” (так называют игровые автоматы с ручкой, дергая за которую получаем случайный выигрыш). Вероятность выиграть на первом автомате $p_1 > 0$, а на втором $p_2 > 0$. Обе вероятности не известны. Игрок может в любом порядке $n \gg 1$ раз дергать за ручки “одноруких бандитов”. Стратегией игрока является выбор ручки на каждом шаге, в зависимости от результатов всех предыдущих шагов, так чтобы суммарный выигрыш был бы максимальным. Приведите асимптотически оптимальную стратегию игрока.

Решите предыдущую задачу, если выигрыш есть случайная величина с распределением из экспоненциального семейства, зависящего от неизвестного параметра θ. Хотя значения θ_1 и θ_2 (для 1-го и 2-го игрового автомата) неизвестны, но, не ограничивая общности, считайте, что $\theta_1 \neq \theta_2$. Выигрышем является сумма выигрышей во всех розыгрышах.

---

[5] *Соколов Г.А., Гладких И.М.* Математическая статистика. – М.: Издательская группа АСТ, 2005.

[6] *Lai T., Robbins H.* Asymptotically efficient adaptive allocation rules // Advances in Applied Mathematics. V. 6. 1985. P. 41–22.

2. (Markov Chain Monte Carlo Revolution.)[7] В руки опытных криптографов попалось закодированное письмо (10 000 символов). Чтобы это письмо прочитать, нужно его декодировать. Для этого берется стохастическая матрица переходных вероятностей $P = \left\| p_{ij} \right\|$ (линейный размер которой определяется числом возможных символов (букв, знаков препинания и т.п.) в языке, на котором до шифрования было написано письмо – этот язык известен и далее будет называться базовым), в которой p_{ij} отвечает за вероятность появления символа с номером j сразу после символа под номером i. Такая матрица может быть идентифицирована с помощью статистического анализ какого-нибудь большого текста, скажем, «Войны и мира» Л. Н. Толстого.

Пускай способ (де)шифрования определяется некоторой, неизвестной, функцией f – преобразование (перестановка) множества кодовых букв во множество символов базового языка.

В качестве «начального приближения» выбирается какая-то функция f, например, полученная исходя из легко осуществимого частотного анализа. Далее рассчитывается вероятность выпадения полученного закодированного текста $\vec{x}$, сгенерированного при заданной функции f (функция правдоподобия):

$$L(\vec{x}; f) = \prod_k p_{f(x_k), f(x_{k+1})} .$$

Случайно выбирается два аргумента у функции f и значения функции при этих аргументах меняются местами. Если в результате получилась такая f^*, что $L(\vec{x}; f^*) \geq L(\vec{x}; f)$, то $f := f^*$, иначе независимо бросается монетка с вероятностью

[7] *Diaconis P.* The Markov chain Monte Carlo revolution // Bulletin (New Series) of the AMS. 2009. V. 49. № 2. P. 179-205. http://www-stat.stanford.edu/~cgates/PERSI/papers/MCMCRev.pdf

выпадения орла $p = L\left(\vec{x}; f^*\right) \Big/ L\left(\vec{x}; f\right)$, и если выпадает орёл, то $f := f^*$, иначе $f := f$. Далее процедура повторяется.

Объясните, почему предложенный алгоритм "сходится" именно к $\overline{f}$? Почему сходимость оказывается такой быстрой (0,01 с на современном PC)?

3. (Оценка вероятности переобучения.)[8] Пусть $x_1,...,x_l$ – простая выборка из распределения с функцией распределения $P(x)$. Элементами этой выборки x_i могут быть, например, векторы. Пусть $\alpha \in \Omega$ – некоторый абстрактный параметр, $0 \le F(x,\alpha) \le a$ – некоторая функция, измеримая при всех $\alpha \in \Omega$ относительно меры $P(x)$. Далее

$$M(\alpha) = MF(x,\alpha) = \int F(x,\alpha)\, dP(x),$$

$$M_{\text{эмп}}(\alpha) = \frac{1}{l}\sum_{i=1}^{l} F(x_i,\alpha).$$

Рассмотрим систему событий S вида $A(\alpha,c) = \{x:\ F(x,\alpha) \ge c\}$ для всевозможных значений $\alpha \in \Omega$ и c. Обозначим $\Delta^S(x_1,...,x_l)$ – число (бинарных) решающих правил класса S, по-разному классифицирующих объекты заданной выборки.[9] Введем функцию роста $M^S(l) = \max \Delta^S(x_1,...,x_l)$, где максимум берется по всем

---

[8] *Червоненкис А.Я.* Компьютерный анализ данных. – М.: Яндекс, 2009. – 260 с.

[9] Выборке $x_1,...,x_l$ и конкретному $A(\alpha,c)$ ставится в соответствие последовательность нулей и единиц по правилу: $x_i \in A(\alpha,c) \Rightarrow 1$, $x_i \notin A(\alpha,c) \Rightarrow 0$. Разным $A(\alpha,c)$ могут соответствовать как разные, так и одинаковые последовательности нулей и единиц. Очевидно, что $\Delta^S(x_1,...,x_l)$ есть число различных последовательностей

последовательностям $(x_1,...,x_l)$ длины l. Покажите, что

$$P\left\{\sup_{\alpha\in\Omega}\left|M(\alpha)-M_{\text{эмп}}(\alpha)\right|>\varepsilon\right\}\le 6M^S(2l)\exp\left[-\frac{\varepsilon^2(l-1)}{4a^2}\right].$$

Замечание.

Заметим, что для любой системы событий S имеет место

$$M^S(l)=2^l \text{ или } M^S(l)\le\sum_{i=0}^{n-1}C_l^i,$$

т.е. $$\exists\ n_0\in\mathbb{N}:\ M^S(l)=\mathrm{O}\left(l^{n_0}\right).$$

Минимально возможное значение n_0 принято называть *размерностью Вапника–Червоненкиса* (VC-размерность). Однако А. Я. Червоненкис предлагает называть её *комбинаторной размерностью S*. Так, например, для множества всевозможных линейных решающих правил в пространстве размерности n комбинаторная размерность равна $n_0=n+1$. Если $M^S(l)=2^l$, то говорят, что комбинаторная размерность бесконечна. Для рассматриваемого в задаче случая достаточным условием конечности комбинаторной размерности, как следствие равномерной сходимости с ростом объема выборки $M_{\text{эмп}}(\alpha)$ к $M(\alpha)$, является условие, что Ω – компакт, $F(x,\alpha)$ непрерывна по α, $|F(x,\alpha)|<K(x)$, где $\int K(x)dx<\infty$.

нулей и единиц, построенных по семейству $\{A(\alpha,c)\}_{\alpha\in\Omega,\,c}$. Очевидно также, что $\Delta^S(x_1,...,x_l)\le 2^l$.

Профессор Андрей Александрович Натан

Данное учебно-методическое пособие основано на концепции курса математической статистики, разработанной профессором А. А. Натаном и развиваемой сейчас его учениками С. А. Гузом, О. Г. Горбачевым, А. В. Гасниковым.

Андрей Александрович НАТАН прожил большую жизнь: родился 6 февраля 1918 г., умер 9 января 2009 г. В 1979–1984 гг. он был деканом факультета управления и прикладной математики (ФУПМ) МФТИ, в 1975 г. участвовал в создании кафедры математических основ управления, которой руководил четверть века. Именно А. А. Натан поставил цикл вероятностно-стохастических дисциплин для студентов ФУПМ, делая основной акцент не только на теоретические конструкции, а и на их прикладное значение.

Особенности личности Андрея Александровича Натана – интеллигентность, скромность, доброжелательность, трудолюбие – ощущались и в его отношении к научной работе. Темы всех его научных исследований продиктованы его гражданской позицией, чувством ответственности за происходящее в стране. Они имеют явную социально значимую прикладную направленность. А. А. Натан создавал и рассматривал математические и имитационные модели, научные результаты как инструменты для решения важных проблем современности. Андрей Александрович долго вынашивал каждую идею, самостоятельно проводил расчеты на компьютере, внимательно и доброжелательно выслушивал и учитывал пожелания и замечания студентов и коллег.

В последние годы, как и в течение всей своей жизни, Андрей Александрович много работал. Результатами этой работы стали оригинальные учебные пособия по теории вероятностей, случайным процессам и математической статистике, подытожившие более чем тридцатилетний опыт чтения лекций и проведения семинаров. В это время Андрей Александрович также подготовил и издал две оригинальные монографии, посвященные стохастическим моделям в микроэкономике и стохастиче-

ским моделям коммерческих операций. Замечательной особенностью этих работ является удивительная ясность, простота предложенных моделей, отражающих и объясняющих тем не менее ряд нетривиальных микроэкономических эффектов и явлений.

Ряд последних работ Андрея Александровича можно найти на сайте http://www.mou.mipt.ru/natan.html

Андрей Александрович охотно и безвозмездно делился идеями, к сожалению, далеко не всегда публикуя их. Задачи, которыми интересовался Андрей Александрович, были актуальными и нетривиальными. Приведем названия некоторых из его последних работ: «Демократия и налог»; «ОСАГО»; «Стохастика и причинность». Особо отметим научно-публицистическую работу «Рассуждения о проблемах и путях развития российского общества». В ней сделана попытка вычленить главные проблемы современного развития российского общества, определить их истоки и взаимосвязи и наметить пути и перспективы решения. Во введении А. А. Натан пишет: «Право автора на такой анализ основано только на его возрасте (1918 год рождения) и на 65-летнем военно-бюрократическом и научно–педагогическом профессиональном опыте в области прикладной математики, а также (главным образом) на его неравнодушии к будущему страны, в которой будут жить его потомки».

ПРИНЯТЫЕ ОБОЗНАЧЕНИЯ

$\mathsf{N}(m,\sigma^2)$ – нормальное распределение с параметрами: m (математическое ожидание) и σ^2 (дисперсия);

$\mathsf{Po}(\lambda)$ – распределение Пуассона с параметром λ;

$\in$ – знак принадлежности к множеству, к типу распределения;

α – вероятность ошибки первого рода, уровень значимости;

β – вероятность ошибки второго рода;

$1-\beta$ – мощность критерия;

MX – математическое ожидание случайной величины X;

DX, σ^2 –дисперсия случайной величины;

$R_{XY} = cov(X,Y)$ – корреляционный момент, ковариация случайных величин X, Y;

r_{XY} – коэффициент корреляции случайных величин X, Y.

$\mathbf{C} = \left(c_{ij}\right)$ – матрица штрафов; c_{ij} – штраф за выбор гипотезы H_j при истинности гипотезы H_i.

с.в. – случайная величина

(у.)з.б.ч. – (усиленный) закон больших чисел

ц.п.т. – центральная предельная теорема

МНП – метод наибольшего правдоподобия

Литература

1. *Натан А.А., Горбачев О.Г., Гуз С.А.* Математическая статистика: учебное пособие. – М.: МЗ Пресс – М.: МФТИ, 2004. – 156 с.
2. *Ивченко Г.И., Медведев Ю.И.* Введение в математическую статистику. – М.: Издательство ЛКИ, 2010. – 600 с.
3. *Лагутин М.Б.* Наглядная математическая статистика. – М.: Бином, 2009. – 472 с.
4. *Боровков А.А.* Математическая статистика. – М.: Физматлит, 2007. – 704 с.
5. *Леман Э.* Проверка статистических гипотез. – М.: Наука, 1979. – 408 с.
6. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975. – 643 с.
7. *Кельберт М.Я., Сухов Ю.М.* Основные понятия теории вероятностей и математической статистики. – М.: МЦНМО, 2007. – 456 с.
8. *Зубков А.М., Севастьянов Б.А., Чистяков В.П.* Сборник задач по теории вероятностей. – М.: Наука, 1989. – 320 с.
9. *Ивченко Г.И., Медведев Ю.И., Чистяков А.В.* Задачи с решениями по математической статистике. – М.: Дрофа, 2007. – 318 с.
10. *Крянев А.В., Лукин Г.В.* Математические методы обработки неопределенных данных. – М.: Физматлит, 2010. – 280 с.
11. *Косарев Е.Л.* Методы обработки экспериментальных данных. – М.: Физматлит, 2008. – 208 с.
12. *Секей Г.* Парадоксы в теории вероятностей и математической статистике. – М. – Ижевск: РХД, 2002. – 240 с.
13. *Воронцов К.В.* Машинное обучение. – М.: МФТИ, 2009. http://www.machinelearning.ru/
14. *Wasserman L.* All of statistics: A concise course in statistic inference. – New York: Springer, 2009. – 442 p.