

# Курс прикладной теории вероятностей и статистики.

Андрей Хохлов

## 1 Цель курса

На мой взгляд, математическая статистика вызывает трудности при изучении прежде всего несоответствием очень математизированного языка теории и необходимостью знания многих полуэвристических приемов обработки данных. Поэтому цель курса я бы сформулировал так: сбалансированно

- сообщить базисные понятия теории с объяснением их возникновения;
- сообщить сведения о языке теории, который в целом очень непрост и достаточно формален;
- изучить хотя бы некоторые методы практической обработки данных доступными на сегодня техническими средствами.

Собранный в курсе список сведений, на мой взгляд, помогает прочесть использующую статистические понятия публикацию, написать конкретные программы обработки данных и делать меньше ошибок при использовании терминологии. Разумеется, курс не претендует на полноту и отбор материала делался в первую очередь для того, чтобы с достигнутых позиций учащемуся можно было бы продвигаться самостоятельно, не путаясь в терминах, методах и понятиях. Некоторые практические задачи по темам (они обычно приведены в начале списка) скорее отвечают названию "лабораторная работа": это достаточно трудоемкие (для того, кто с ними сталкивается впервые) действия, которые надо реализовать доступными программными средствами, например, на MATLAB. Они должны быть разобраны вместе с преподавателем на семинарских занятиях и не предполагают самостоятельного решения студентом. В целом, список задач построен так, чтобы стимулировать самостоятельное прочтение материала из рекомендуемых книг.

## 2 Практические действия с выборками и вероятностные описания.

Данные об измерениях и построение вероятностных моделей. Одномерные и многомерные распределения, статистические зависимости.

### 2.1 Литература к теме.

Гнеденко "Теория вероятностей"

Тутубалин "Теория вероятностей и случайных процессов"

### 2.2 Задачи к теме.

1. Методом Монте-Карло оценить вероятность попадания направлений в заданный телесный угол. Случайные направления определяются началом координат и трехмерным гауссовским распределением общего вида.
2. Для **равномерно распределенной** на отрезке  $[a, b]$  случайной величины  $\alpha$  аналитически найти формулы **плотностей распределения** случайных величин  $|\alpha|$ ,  $\alpha^2$ ,  $\sqrt{|\alpha|}$ .
3. Для стандартной равномерной на отрезке  $[0, 1]$  случайной величины  $\alpha$  найти распределение  $-\log_5 \alpha$ .
4. В области  $\mathcal{G}$  (например,  $\mathcal{G} = \{|x| + |y| < 1\}$ ) задана двумерная плотность распределения

$$f_{(\alpha, \beta)}(x, y) = \begin{cases} \text{const} & x, y \in \mathcal{G} \\ 0 & x, y \notin \mathcal{G} \end{cases}$$

- Найти плотности  $f_\alpha(x)$   $f_\beta(y)$  распределения значений первой  $\alpha$  и второй  $\beta$  координат вектора.
  - Зависимы ли  $\alpha$  и  $\beta$ ?
5. Найти явную формулу для плотности распределения случайной величины  $\xi = \alpha^2 + \beta$ , где  $\alpha$  и  $\beta$  определены в предыдущей задаче.
  6. Плотность распределения случайной величины  $\gamma$  выражается формулой  $C \cdot e^{b|x-a|}$  где параметры  $C, b$  известны. Найти явные выражения для оценки величины  $a$  методом максимального правдоподобия и методом наименьших квадратов.
  7. Для случайной величины  $\gamma$  из предыдущей задачи доказать неравенство Гаусса

$$P\left\{|\gamma - a| \geq \varepsilon \sqrt{E(|\gamma - a|^2)}\right\} \leq \frac{4}{9\varepsilon^2}$$

8. Установить из серии следующих задач что неравенство Гаусса справедливо для любой случайной величины, плотность которой максимальна в  $a$  и убывает с ростом  $|x - a|$ .
  - Найти максимальную площадь прямоугольника  $\Delta x \times \Delta y$ , который можно вписать в криволинейный треугольник  $0 \leq x \leq 1, 0 \leq y \leq x^2$ .
  - Доказать, что для любого  $0 < u < N$  и всюду невозрастающей на положительной полуоси неотрицательной функции  $f(x)$  выполнено  $\int_0^N x^2 f(x) dx \geq \frac{9}{4} u^2 \int_u^N f(x) dx$ . (Подсказка: рассмотреть приближения функции  $f(x)$  кусочно-постоянными)
  - Если несобственные интегралы существуют, то для  $0 < u$  выполнено  $\int_0^\infty x^2 f(x) dx \geq \frac{9u^2}{4} \int_u^\infty f(x) dx$ . Доказать.
  - Положим  $d^2 = E(|\gamma - a|^2)$  и запишем  $P\{|\gamma - a| \geq \varepsilon d\}$  через плотность распределения случайной величины  $\gamma$ . Используя неравенство предыдущей задачи получить неравенство Гаусса.

### 3 Основные статистические конструкции и принципы.

Свертки, корреляции, копулы. Моменты, семиинварианты, производящие и характеристические функции. В каком смысле можно восстановить распределение по его характеристикам? Тяжелые хвосты распределений. Несуществование моментов, характеристики распределений с тяжелыми хвостами. Что означают равенства и предельные переходы для случайных величин? Практический смысл и границы применимости ЦПТ в приложениях. Парадоксальные варианты статистических оценок: парадокс Байеса, парадокс оценок мат.ожидания, парадокс оценок дисперсии.

#### 3.1 Литература к теме.

Ширяев "Вероятность".

Тутубалин "Теория вероятностей и случайных процессов"

Roger, Nelsen "An introduction to copulas".

Секей "Парадоксы в теории вероятностей и мат.статистике".

Press, Teukolsky, et al "Numerical Recipes"

#### 3.2 Задачи к теме.

1. По заданной выборке (достаточно большого объема) значений одномерной случайной величины  $\gamma$  выяснить аргументы в пользу гипотезы существования или несуществования момента  $m_k(\gamma)$  степени  $k$ .
2. Проверить для заданной выборки одномерной случайной величины гипотезу о форме распределения методом Колмогорова-Смирнова и методом Адельсона-Дарлингга.
3. Проверить ассоциативность и коммутативность для копулы  $C_\alpha(x, y) = xy[1 + \alpha(1-x)(1-y)]$ ,  $-1 \leq \alpha \leq 1$ ,  $\alpha \neq 0$ .

- Пусть случайная величина  $X$  равномерно распределена на интервале  $(-1, 1)$  и  $Y = |X|$ . Очевидно, что между  $X$  и  $Y$  имеется функциональная связь. Вычислить коэффициент корреляции  $\rho(X, Y)$  и установить коррелированы ли эти случайные величины.
- Для случайного вектора размерности большей двух попарные независимости координат вовсе не гарантируют независимость по тройкам. Построить соответствующий пример.
- Практические интерпретации формул. Пусть имеются два разных карандаша, рассмотрим два способа определения их длин. В первом случае померяем каждый карандаш по отдельности. Во втором - непосредственно измерим (в этом случае прикладывая карандаши один к другому) суммарную длину и разность длин и выразим длины из системы уравнений. Считая точность всех измерений одинаковой сравнить получившиеся среднеквадратичные ошибки. При каком способе среднеквадратическая ошибка будет меньше?

## 4 Случайное блуждание и связанные с ним задачи

Переходные вероятности, первые достижения, время до разорения – в терминах производящих функций. Гармонические функции и потенциалы. Случайные блуждания на прямой, плоскости и в пространстве. Переход к пределу (диффузия). Марковские цепи. Понятие случайного процесса, основные примеры: винеровский процесс, пуассоновский процесс. Многомерное нормальное распределение. Комплекснозначные гауссовские процессы: модели дискретного и непрерывного времени. Стохастические модели для случайных точек. Пуассоновский процесс на прямой и в многомерных пространствах. Теоремы о суперпозиции и теорема об отображении. Обратная и прямая системы дифференциальных уравнений Колмогорова. Парадоксы необратимости марковских процессов.

### 4.1 Литература к теме

Феллер "Введение в теорию вероятностей и ее приложения" том 1, 2

Дынкин, Юшкевич "Теоремы и задачи о процессах Маркова"

Кингман "Пуассоновские процессы"

### 4.2 Задачи к теме

- По заданному временному ряду провести оценивание его ковариационной функции.
- По заданному ансамблю реализаций дробного броуновского движения оценить показатель Гельдера и соответствующий параметр в процессе. Сравнить с оценками вдоль траекторий.
- По ковариации и среднему на основе генератора стандартной гауссовой случайной величины построить реализации (в дискретном времени) гауссова стационарного процесса с заданными характеристиками.
- Рассмотрим дискретное случайное блуждание, начинающееся из точки  $x$ , и обозначим переходную функцию блуждания через  $p(n, x, y)$  (то есть вероятность того, что на  $n$ -м шаге мы попадаем в точку  $y$ :  $x(n) = y$ ). Тогда сумма

$$g(x, y) = \sum_{n=0}^{\infty} p(n, x, y)$$

(то есть дискретный потенциал) представляет собой математическое ожидание числа попаданий в точку  $y$ . Доказать.

- При симметричном случайном блуждании по двумерной решетке  $g(x, x) = \infty$ . Доказать.
- В симметричном случайном блуждании по двумерной решетке обозначим  $r(x)$  вероятность того, что вышедшая из точки  $x$  частица когда-либо вернется в  $x$ . Показать, что

$$g(x, x) = \sum_{n=0}^{\infty} r^n(x)$$

## 5 Язык теории случайных процессов

Предмет теории случайных процессов, некоторые задачи. Предел измеримых отображений. Построение семейства независимых случайных элементов с заданными распределениями. Процессы частных сумм, процессы восстановления, модель страхования Крамера-Лундберга, пуассоновская случайная мера. Случайная функция как семейство случайных элементов и как одно измеримое отображение. Конечномерные распределения случайной функции. Процессы с непрерывными траекториями. Согласованность проекций меры. Эквивалентные случайные функции. Стационарные в широком смысле процессы и их ковариационные функции.

### 5.1 Литература к теме

Булинский, Ширяев "Лекции по теории случайных процессов".

Karatzas, Shreve "Brownian Motion and Stochastic Calculus"

Булинский, Ширяев "Лекции по теории случайных процессов".

### 5.2 Задачи к теме

1. Докажите, что на  $([0, 1], \mathfrak{B}([0, 1]), P)$  где  $P$  — мера Лебега, невозможно построить континуальное семейство  $\{\xi_t, t \in \mathbb{R}\}$  независимых случайных величин чтобы  $P(\xi_k = 1) = P(\xi_k = 0) = 1/2$  при каждом  $t$ .
2. Рассмотрим на некотором вероятностном пространстве  $(\Omega, \mathfrak{F}, P)$  две последовательности неотрицательных случайных величин  $\{\xi_k, k \in \mathbb{N}\}$  и  $\{\eta_k, k \in \mathbb{N}\}$ . При этом все величины внутри каждой последовательности имеют одинаковые распределения, а также  $\{\xi_k, \eta_k, k \in \mathbb{N}\}$  независимы в совокупности. Для положительных констант  $c, y_0$  определим процессы, принимающие значения в  $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$

$$X_t(\omega) = \sup \left\{ n : \sum_{j \leq n} \xi_j(\omega) \leq t \right\}, X_0(\omega) = 0, t > 0$$
$$Y_t(\omega) = y_0 + ct - \sum_{j=1}^{X_t(\omega)} \eta_j(\omega), t \geq 0$$

Процессы  $Y_t$  используются в модели страхования Крамера-Лундберга, в которой  $y_0$  — начальный капитал компании,  $c$  — скорость поступления взносов,  $\eta_j$  — размер выплаты в случайный момент  $\tau_j = \sum_{i=1}^j \xi_i$ ,  $j \in \mathbb{N}$ ,  $Y_t$  интерпретируется как капитал компании в момент  $t$ . Нарисовать вид типичной траектории процесса  $Y_t$ .

3. Пусть  $\{\xi_k, k \in \mathbb{N}\}$  — последовательность независимых одинаково распределенных случайных величин, заданных на вероятностном пространстве  $(\Omega, \mathfrak{F}, P)$ . Введем эмпирические меры

$$P_n(B, \omega) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_B(\xi_j(\omega)), B \in \mathfrak{B}, n \in \mathbb{N}$$

где, как обычно,  $\mathbf{1}_B$  индикатор множества  $B$ ,

$$\mathbf{1}_B(x) = \begin{cases} 1, & x \in B \\ 0, & x \notin B \end{cases}$$

Положим  $F_n^*(x, \omega) = P_n((-\infty, x], \omega)$ ,  $x \in \mathbb{R}$  этот процесс принято называть эмпирической функцией распределения. Постройте график траектории процесса  $F_n^*(x, \omega)$ , в чем будет отличие случая, когда  $\xi_j$  равномерно распределены на  $[0, 1]$  от случая, когда  $\xi_j$  — бернуллиевские величины с  $P(\xi_k = 1) = 1 - q$ ,  $P(\xi_k = 0) = q$  (здесь  $q > 0$  — параметр)?

## 6 Стохастический интеграл

Ортогональные случайные меры и их построение. Интеграл по ортогональной случайной мере, его свойства. Теорема Карунена-Хинчина о факторизации ковариационной функции и представлении процесса в виде интеграла по ортогональной случайной мере. Интеграл для случайных функций по винеровскому процессу. Интеграл Стратоновича. Конструкция Ито стохастического интеграла для неупреждающих случайных функций.

Свойства стохастического интеграла. Формула замены переменных Ито. Стохастические дифференциальные уравнения. Уравнение Ланжевена. Уравнение Орнштейна-Уленбека. Уравнение Фоккера-Планка (одномерное) как прямое уравнение Колмогорова.

## 6.1 Литература к теме

Karatzas, Shreve "Brownian Motion and Stochastic Calculus"

Булинский, Ширяев "Лекции по теории случайных процессов".

Risken H. "The Fokker-Planck equation. methods of solution and applications"

Johnson, Lapidus "The Feynman Integral and Feynman Operational Calculus"

## 6.2 Задачи к теме

1. Следуя главе 3.6 "Solutions of the Langevin Equation by Computer Simulation" из книги Risken H. "The Fokker-Planck equation. methods of solution and applications" выполнить построение решения уравнения Ланжевена.
2. Вычислить интеграл по винеровской мере  $\mathbf{m}$  на пространстве путей  $\mathbf{x}(t) \in C_0^{A,B}$  ( т.е. путей  $\mathbf{x}(t)$  на отрезке  $A, B$ , выходящих из 0) от функции  $F : C_0^{A,B} \rightarrow \mathbb{R}$  заданной формулой

$$F(\mathbf{x}) = \|\mathbf{x}\|^2 = \int_A^B \mathbf{x}^2(t) dt$$

(указание: см. Главу 3 в книге Johnson, Lapidus "The Feynman Integral and Feynman Operational Calculus").

3. Для  $A < t_1 < t_2 < t_3 \leq B$  вычислить явно интеграл ( по винеровской мере  $\mathbf{m}$  на том же пространстве путей)

$$\int_{C_0^{A,B}} \mathbf{x}(t_1)\mathbf{x}(t_2)\mathbf{x}(t_3) d\mathbf{m}$$

4. Вычислить ( по винеровской мере  $\mathbf{m}$  на том же пространстве путей) интеграл от функции  $F : C_0^{A,B} \rightarrow \mathbb{R}$  заданной формулой

$$F(\mathbf{x}) = \left( \int_A^B \mathbf{x}(t) dt \right)^2$$