Компьютерная лингвистика, 2019-20 уч. год
Мат. основания комп. лингвистики (http://math-info.hse.ru/s19/a)
Семинар 1 (16 сентября 2019)

И. Щуров, Д. Леонкин

Задача 1. Случайная величина X принимает три значения: 1, 2 или 3 и имеет распределение, зависящее от параметра $\theta \in [0, 1/2]$.

x	1	2	3
$P(X = x)$	θ	θ	$1 - 2\theta$

Пусть выборка
(a) $3, 1, 2$;
(b) $1, 2, 2, 3, 3, 3$

получена как результат последовательных независимых реализаций случайной величины X. Найти оценку максимального правдоподобия для параметра θ.

Задача 2. Пуассоновская случайная величина принимает любые целые неотрицательные значения и её распределение задаётся следующим образом:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

где $\lambda > 0$ — параметр распределения, $k = 0, 1, 2, \ldots$

Пусть выборка $1, 5, 3$ получена как результат последовательных независимых реализаций пуассоновской случайной величины с параметром λ. Найти оценку наибольшего правдоподобия для λ.

Задача 3. Геометрически распределённая случайная величина принимает любые натуральные значения. Её распределение задаётся следующим образом:

$$P(X = k) = (1 - p)^{k-1} p,$$

где $p \in (0, 1)$ — параметр распределения, $k = 1, 2, \ldots$

Пусть выборка $1, 3, 6$ получена как результат последовательных независимых реализаций геометрической случайной величины с параметром p. Найти оценку наибольшего правдоподобия для p.

Замечание 1. *Если случайная величина не является дискретной, в качестве правдоподобия выборки используется не вероятнось её получения, а плотность вероятности.*

Задача 4. Рассмотрим непрерывную случайную величину X с плотностью

$$p(x) = \begin{cases} 2 - 2x, & x \in [0, 1]; \\ 0, & \text{иначе.} \end{cases}$$

Пусть $Y_\theta = X + \theta$, где θ — некоторый параметр.

(a) Найти плотность Y_θ.
(b) Рассмотрим выборку из Y_θ, состоящую из единственного числа 3. Найти оценку максимума правдоподобия для θ.

Задача 5. Решить предыдущую задачу, если случайная величина X имеет плотность

$$p(x) = \begin{cases} x, & x \in [0,1]; \\ 2-x, & x \in [1,2]; \\ 0, & \text{иначе.} \end{cases}$$

Задача 6. Нормально распределённая случайная величина принимает произвольные вещественные значения. Её функция распределения задаётся следующим образом:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Пусть выборка
(a) 7;
(b) 1, 3, 7, 2, 5

получена как результат независимых реализаций нормально распределённой случайной величины с неизвестным параметром μ и известным $\sigma = 1$. Найти оценку наибольшего правдоподобия для μ. Изменится ли она, если σ окажется равно другому числу?

Задача 7. Построить график функции распредедления для случайной величины X из задачи 1 при $\theta = 0{,}2$.

Задача 8. Равномерно распределённая случайная величина на отрезке $[0,1]$ имеет плотность

$$p(x) = \begin{cases} 1, & x \in [0,1]; \\ 0, & \text{иначе.} \end{cases}$$

Построить график функции распределения для такой случайной величины.

Задача 9. Рассмотрим выборку $(x_1, x_2, \ldots, x_n)$, $x_i = i/n$, $i = 1, \ldots, n$. Для
(a) $n = 2$;
(b) $n = 5$;
(c) $n = 10$

построить график эмпирической функции распредедления для данной выборки. Сравнить с графиком из предыдущей задачи.