МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ Н. Э. БАУМАНА

Факультет информатики и систем управления

Кафедра теоретической информатики и компьютерных технологий

Лекции по курсу «Математическое моделирование»

Выполнил:
студент ИУ9-111
Выборнов А. И.

Лекции читала:
Домрачева А.Б.

Москва 2015

Содержание

1. 2015-09-29

Изучением математических моделей случайных явлений или экспериментов в первую очередь занимаются такие науки, как мат. статистика (МС) и теория вероятности (ТВ).

Задачи МС являются обратными к задачам ТВ. В ТВ после задания того или иного случайного явления требуется расчитать вероятностные характеристики в рамках данной модели. Моделирование производится на основе результата эксперимента называемых статистическими данными. В ряде случаев по результатам эксперимента требуется лишь уточнить или модифицировать имеющуюся модель. В задачах МС вероятность того или иного события известна и необходимо оценить параметры эксперимента (параметры функции связи между двумя показателями объекта, параметры закона распределения случайной величины, в более широком случае функцию распределения случайной величины или функцию плостности распределения случайной величины).

1.1. Основные задачи мат. статистики

- **Задача оценки неизвестных параметров по результатам эксперимента.** Как правило нужно найти функцию от результатов эксперимента, является достаточно хорошей оценкой неизвестного истинного значения параметра (a — параметр, $\hat{a}$ — оценка параметра).

- **Задача интервального оценивания.** Есть строгий интервал со случайными границами (нижняя — $a\_$, верхняя — $a-$), таким образом, чтобы он накрывал неизвестное истинное значение параметра с заранее заданной веростностью γ.

$$P\{a_- \leq a \leq a^-\} = \gamma$$

.

- **Задачи проверки статистических гипотез.** Требуется, на основе математических экспериментов, проверить то, или иное предположение относительно вида, и параметра функции распределения случайной величины, и функции плотности распределения случайной величины.

В мат.статистике используется выборочная терминология основанная на ”урновой" схеме. Пусть имеется урна содержащая N чисел

$$\{X_1, X_2, ..., X_N\},\ (1)$$

называемая генеральной совокупностью объёмом N. Набор 1 может иметь бесконечную размерность. Из генеральной совокупности выбирается набор

$$\{x_1, x_2, ..., x_n\}, n \leq N.\ (2)$$

Набор 2 называется выборкой объёма n из генеральной совокупности 1.

Выборка может производится с возвращением и без возвращения. Если выборка производится с возвращением, то случайные величины в ней независимы. С возвращением это независимая, повторная, случайная выборка объёмом n. Терминология сохраняется и в случае бесконесной генеральной совокупности.

Числа выборки 2 обычно располагают в порядке убывания или возврастания:

$$\{x^{(1)}, x^{(2)}, ..., x^{(n)}\}.\ (3)$$

Набор 3 называется вариационным рядом. Чаще всего в задачах это называется вариационный ряд.

Эмпирической функцией распределения построенной на основе выборки 3 называется функция $\hat{F}(x) = \frac{r(x)}{n}$ (n — общее число выборки, $r(x)$ — количество элементов выборки $x_i \leq x$).

Пример: Выборка $0, 0, 9, 16, 21, 24, 29, 37, 42, 48$.

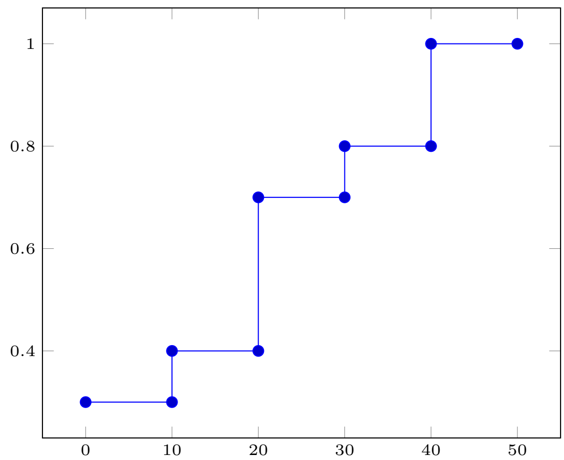

Рисунок 1 — График $\hat{F}(x)$

Для моделирования требуется теоретическая функция распределения случайной величины x. Которая может быть оценена по эмпирической функции распределения.

По теореме гливенко-кантелли: $\sup_{x,n\to\infty} |F(x) - \hat{F}_h(x)| \to 0$

По эмпирической функции распредления строят ... модели.

Выборочное среднее (эмперическое среднее) - $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Выборочный аналог первого начального момента (мат.ожидания).

$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})$ - выборочнаяя (эмпирическая) дисперсия

$\hat{S} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})}$ выборочное СКО (средне квадратичное отклонение)

$\mu_{r,a} = (\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^r)^{\frac{1}{r}}$ - выборочные моменты порядка r.

В ряде случаев требуется оценить размах выборки $R_n = |x^{(n)} - x^{(1)}|$.

1.2. Точечные оценки параметров

Пусть имеется некоторая случайная величина ξ с функцией распределения $F(x,\theta)$, плотностью распределения $f(x,\theta)$.

Обычно говорят о параметрическом семействе распределений, в котором θ принимает различные значения.

Вводят функцию от результатов наблюдений

$$\phi = \phi(x_1, x_2, ..., x_n) \ (4)$$

(x_i - элемент набора 2), называемую статистикой. Задача построения точечной оценки параметра θ, сводится к нахождению значения статистики. Такой что

$$\hat{\theta} = \theta(x_1, x_2, ..., x_n) : \sup_{n\to\infty} |\hat{\theta} - \theta| \to 0$$

Необходимо установить эффективную оценку, рекомендуемую в качестве результата.

1.2.1. Свойства оценок

$\hat{\theta} = \theta(x_1, x_2, ..., x_n)$

Пример: $\hat{\lambda} = 1/\overline{x} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{n}{x_1+x_2+...+x_n}$

$\hat{\lambda} = \lambda(x_1, x_2, ..., x_n) = \frac{n}{x_1+x_2+...+x_n}$

$M\hat{\theta} = \theta$

Оценка $\hat{\theta}$ является несмещённой оценкой параметра θ, если её мат. ожидание совпадает с теоритической величиной.

$\lim_{n\to\infty} M\hat{\theta_n} = \theta$

Таким образом на свойство оценок влияет объём выборки.

Если $\lim_{n\to\infty} P\{|\hat{\theta_n} - \theta| < \epsilon\} \to 1$ мы говорим о состоятельности оценок. Сходится по вероятности.

Пусть $\hat{\theta_n}$ асимптотически несмещённая оценка параметра θ. В случае когда $\lim_{n\to\infty} S^2(\hat{\theta_n}) \to 0$ оценка состоятельна.

Таким образом асимптотическая несмещённость оценки θ и минимизация разброса значений параметра при $n \to \infty$ обеспечивают состоятельность оценки (теорема приводится без доказательства).

Пусть имеются две оценки $\hat{\theta_n}^1$, $\hat{\theta_n}^2$.

$S^2(\hat{\theta_n}^1) = M(\hat{\theta_n}^1 - \theta)^2 \leq M(\hat{\theta_n}^2 - \theta)^2 = S^2(\hat{\theta_n}^2)$

То оценка $S^2(\hat{\theta_n}^1)$ является более эффективной по сравнению с $S^2(\hat{\theta_n}^2)$.