

Министерство образования Иркутской области

Государственное бюджетное профессиональное образовательное

учреждение Иркутской области

“Иркутский авиационный техникум”

(ГБПОУИО “ИАТ”)

Вероятно-статистический подход к измерению информации.

Практико-ориентированный проект по дисциплине ПОД.14

Математика

РУКОВОДИТЕЛЬ: (К.Н. ИЛЬИНЕЦ)

ОБУЧАЮЩИЙСЯ: (К.С.МЯСНИКОВ)

Введение

Цели проекта: цель моего проекта узнать кто и каким образом открыли вероятность статистический подход к измерению информации а также изучить понятие обработки информации как таковой

Задачи проекта:

- Выяснить, каким образом работает подобный подход к измерению информации, как им пользоваться и зачем вообще он нужен.
- Выяснить какие этапы проходит информация и что с ней происходит.

Актуальность

- ▶ В наше время люди много работают с огромными массивами информации и что бы с ними было проще работать люди пользуются программами для обработки разного вида информации, это значительно упрощает жизнь и экономит время. Пример такой программы-тот же самый MS Excel

История появления метода

В 30-х годах XX века американский ученый Клод Шеннон предложил связать количество информации, которое несет в себе некоторое сообщение, с вероятностью получения этого сообщения.

$$I = -\sum_{i=1}^N p_i \log_2 p_i$$

Вероятностный подход к измерению информации

- ▶ Определить понятие "количество информации" довольно сложно. В решении этой проблемы существуют два основных подхода. Исторически они возникли почти одновременно. В конце 40-х годов XX века один из основоположников кибернетики американский математик Клод Шеннон развил вероятностный подход к измерению количества информации, а работы по созданию ЭВМ привели к "объемному" подходу.

Этапы обработки информации.

На самом деле их там много, но я выделил самые нужные.

1. Этапы обработки:
2. Сбор информации
3. Группировка данных
4. Составление таблицы данных
5. График распределения данных
6. Паспорт данных

Это все этапы, которые нам нужны, с ними разберемся далее



Характеристики обработки информации

Размах R - разность между наибольшей и наименьшей вариантой.

$$R = x_{\max} - x_{\min}$$

Мода M_o - варианта которая встречается больше всего раз.

Среднее значение M - среднее значение всех вариантов

$$M = \frac{Xt_1 + Xt_2 + Xt_n}{n}$$

Дисперсия D - мера разброса значений случайной величины относительно её математического ожидания

$$D = \frac{Xt_1 \times S_1 + Xt_2 \times S_2 + Xt_n \times S_n}{n}$$

Частота V - величина, характеристика периодического процесса, равна количеству повторений или возникновения событий (процессов) в единицу времени. Рассчитывается, как отношение количества повторений или возникновения событий (процессов) к промежутку времени, за которое они совершены отношение кратности (сколько раз информация встретилась) на ее объем. частота=кратность деленная на объем, частота%= частота * 100%

$$v = \frac{n}{t}$$

Характеристики обработки информации

Отклонение-
отклонение варианты
от среднего значения
 $x_i - M$

Квадрат
отклонения-
отклонение
квдрате Σ
 $(x_i - M)^2$

Квадратный корень из
дисперсии,
называется
среднеквадратическ
им отклонением,
стандартным
отклонением
или стандартным
разбросом.

$$\delta = \sqrt{D}$$

Сбор информации

- Рассмотрим пример: В некотором классе было 25 учеников, им на каникулы задали прочитать 6 книг, после каникул у них спросили сколько книг они прочитали в результате получили следующие значения.

Варианта X_i	0	1	2	3	4	5	6	Всего :7
Кратность	1	2	4	7	5	4	2	n=25
Частота	0.04	0.08	0.16	0.28	0.20	0.16	0.08	1
Частота %	4	8	16	28	20	16	8	100%
Средне M	3.32							
Отклонение S	-3.32	-2.32	-1.32	-0.32	0.68	1.68	2.68	
Квадрат отклонения S^2	11.02	5.38	1.74	0.10	0.46	2.82	7.18	
Дисперсия D	2.29							
Средн.квад. отклонение	1.51							

Размах: $R=6-0=6$

Мода: $M_o = 3$

Среднее: $M=(0*1+1*2+2*4+3*7+4*5+4*4+6*2) : 25=3.32$

Дисперсия: $D=(0*11,02+1*5,38+2*1,74+3*0,10+4*0,46+4*2,82+6*7,18):25=2,297$

Составление графика распределения данных

Группировка данных и составление таблицы данных, которые я объединил в один этап, чтобы не растягивать:

На основе полученных данных мы можем составить паспорт данных

1. $N=25$
2. $\delta = 1.51$
3. $D=2,29$
4. $R=6$
5. $M=3.32$
6. $Mo = 3$

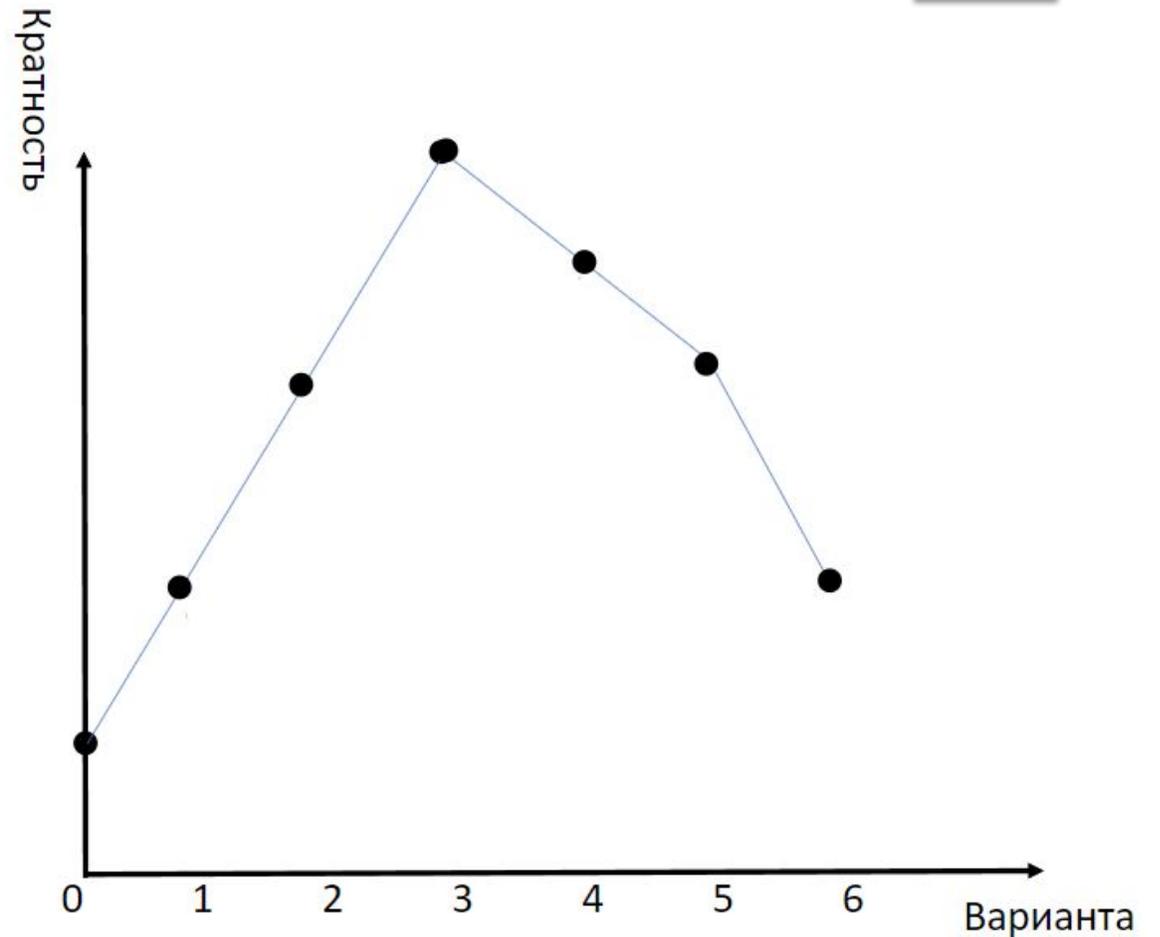


График распределения данных

Виды обработки информации

В информатике, как правило, измерению подвергается информация, представленная дискретным сигналом. При этом различают следующие подходы:

- Семантический подход. Учитывает целесообразность и полезность информации. Применяется при оценке получаемой информации.
- Структурный подход. Измеряет количество информации простым подсчетом информационных элементов, составляющих сообщение.
- Статистический подход. Учитывает вероятность появления сообщений: более информативным считается то сообщение, которое менее вероятно, т.е. менее всего ожидалось. Применяется при оценке значимости получаемой информации.

$$I = \log_2 \frac{P_1}{P_2},$$

$$H = \log_2 (N)$$

$$I = -\log_2 p,$$

Статистический подход к измерению информации

- Статистический подход основывается на учете вероятности появления сообщений. Считается, что сообщение более информативное, если оно менее вероятно (ожидаемо). Данный подход применяется для вычисления значимости получаемой информации.

$$I = -\log_2 p,$$

Семантический подход к измерению информации

- Семантический подход обеспечивает учет целесообразности и полезности информации. Он используется для вычисления эффективности получаемой информации и ее соответствия реальности. Семантический метод включает такие меры измерения информации как: целесообразность, полезность и истинность информации. Целесообразность информации. Количество информации I с позиций ее истинности, получаемой вместе с сообщением, определяется по формуле

$$I = \log_2 \frac{P_1}{P_2},$$



Аддитивная (логарифмическая) мера (структурный подход)

Определяет максимально возможное количество информации в заданных объемах. Единица измерения – информационный элемент. Мера может быть использована для определения информационной емкости памяти компьютера. В этом случае в качестве информационного элемента выступает минимальная единица хранения – бит.

$$H = \log_2(N)$$

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

- YouTube канал: [Видеокурсы DA VINCI - YouTube](#)
- YouTube канал: [Dmitry Kivganov - YouTube](#)
- YouTube канал: [ИНФОУРОК - YouTube](#)
- Сайт в интернете: [Монография обработка данных Калинин АГ 2014.pdf \(sibupk.su\)](#)
- Информатика. Базовый курс. / Под ред. С.В.Симоновича. — Спб., 2000 г.
- Информатика. Компьютерная техника. Компьютерные технологии. / Пособие под ред. О.И.Пушкаря.— Издательский центр "Академия", Киев, — 2001 г.
- Казиев В.М. Введение в анализ, синтез и моделирование систем. М.: Бином, 2007.
- Коцюбинский А.О., Грошев С.В. Современный самоучитель профессиональной работы на компьютере. — Г.: Триумф, 1999 г.