

ПРЕДИСЛОВИЕ

Как и следует из названия данной книги, она посвящена байесовской статистике.

Данная книга может оказаться полезной:

- тем, кто начинает изучать байесовскую статистику;
- тем, кто затрудняется в понимании различий между обычной статистикой и байесовской;
- тем, кто планирует поступать на математические факультеты.

Чтобы не испытывать затруднений во время чтения данной книги, необходимо обладать знаниями математики на уровне старшей школы. Также без знания так называемой обычной, или традиционной, статистики разобраться в данной книге будет непросто.

Книга состоит из следующих глав:

- 1 «Что такое байесовская статистика?»;
- 2 «Базовая информация»;
- 3 «Функция правдоподобия»;
- 4 «Теорема Байеса»;
- 5 «Методы Монте-Карло для цепей Маркова»;
- 6 «Примеры практического использования методов Монте-Карло для цепей Маркова».

До четвертой главы включительно материал в книге несложный. Но начиная с главы 5 уровень сложности возрастает.

Каждая глава при этом состоит из:

- раздела манги;
- дополнительного текстового раздела.

Правда, есть и главы без дополнительного раздела. В данной книге можно читать только разделы с мангой. Пропуск дополнительного материала не мешает пониманию следующих глав.

В этой книге приводятся довольно подробные вычисления. Читателям, склонным к математике, рекомендуется их внимательно изучить. Если же вы не очень дружите с математикой или у вас мало времени, то достаточно быстро просмотреть приводимые вычисления, чтобы составить общее впечатление о том, как их следует производить. Нет необходимости сразу как следует разбираться в сложных вычислениях. Но всё-таки непременно просмотрите их.

Я хочу выразить благодарность издательству Ohmsha за предоставленную возможность написать данную книгу. В особенности я благодарен господину Цукуи Ясухико, который в данной компании занимается моими книгами, начиная с самой первой «Разбираемся с помощью манги. Статистика». А также выражаю признательность господину Уэдзи Юхо и АО Verte, которые помогли превратить мой текст в мангу.

Ноябрь 2017 года

Такахаси Макото

СОДЕРЖАНИЕ

Предисловие	V
Пролог. Хочу изучать байесовскую статистику!	1
Глава 1. ЧТО ТАКОЕ БАЙЕСОВСКАЯ СТАТИСТИКА? ..	11
1.1. Байесовская статистика.....	12
1.2. Различие между обычной и байесовской статистиками	18
Глава 2. БАЗОВАЯ ИНФОРМАЦИЯ	25
2.1. Математическое ожидание, дисперсия, среднеквадратическое отклонение	29
1.1.1. Математическое ожидание	29
1.1.2. Дисперсия и среднеквадратическое отклонение.....	30
2.2. Вероятностное распределение.....	32
2.2.1. Равномерное распределение.....	33
2.2.2. Биномиальное распределение	34
2.2.3. Мультиномиальное распределение.....	40
2.2.4. Равномерное распределение	48
2.2.5. Нормальное распределение.....	49
2.2.6. t -распределение	50
2.2.7. Обратное гамма-распределение	51
2.3. Прочие вероятностные распределения	55
2.3.1. Отрицательное биномиальное распределение.....	55
2.3.2. Распределение Пуассона.....	57
2.3.3. Экспоненциальное распределение.....	60
2.3.4. Бета-распределение.....	62

Глава 3. ФУНКЦИЯ ПРАВДОПОДОБИЯ	63
3.1. Правдоподобие	68
3.1.1. Закон больших чисел	68
3.1.2. Информационное расхождение Кульбака–Лейблера.....	72
3.1.3. Правдоподобие	77
3.2. Функции правдоподобия.....	79
3.2.1. Функция правдоподобия мультиномиального распределения	79
3.2.2. Функция правдоподобия нормального распределения	85
3.3. Другие функции правдоподобия.....	93
3.3.1. Функция правдоподобия биномиального распределения	93
3.3.2. Функция правдоподобия распределения Пуассона.....	95
 Глава 4. ТЕОРЕМА БАЙЕСА	 97
4.1. Теорема Байеса.....	102
4.1.1. Условная вероятность	102
4.1.2. Одновременная вероятность.....	105
4.1.3. Теорема Байеса	106
4.1.4. Показательные примеры.....	107
4.2. Априорная и апостериорная функции плотности вероятности	112
 Глава 5. МЕТОДЫ МОНТЕ-КАРЛО ДЛЯ ЦЕПЕЙ МАРКОВА	 117
5.1. Интегрирование Монте-Карло.....	124
5.1.1. Интегрирование Монте-Карло.....	124
5.1.2. Математическое ожидание и дисперсия непрерывных случайных величин	127
5.2. Цепи Маркова	130
5.2.1. Цепи Маркова	130
5.2.2. Инвариантное распределение	132
5.3. Методы Монте-Карло для цепей Маркова	136

5.3.1. Методы Монте-Карло для цепей Маркова.....	136
5.3.2. Алгоритм Метрополиса–Гастингса	139
5.3.3. Семплирование по Гиббсу.....	156
5.4. Естественное сопряжённое априорное распределение.....	172

Глава 6. ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ МЕТОДОВ МОНТЕ-КАРЛО ДЛЯ ЦЕПЕЙ МАРКОВА 175

6.1. Предположение о среднем значении в двух генеральных совокупностях.....	176
6.1.1. Проверка статистических гипотез.....	178
6.1.2. Процедура проверки статистических гипотез.....	178
6.1.3. Виды проверок статистических гипотез, нулевая и альтернативная гипотезы.....	180
6.1.4. Пример.....	181
6.2. Байесовское иерархическое моделирование	186

ПРИЛОЖЕНИЕ 207

1. Предпосылки априорного распределения и апостериорное распределение	208
1.1. Type B.....	208
1.2. Type C	212
1.3. Подведение итогов	214
2. Проверка сходимости	215
2.1. Метод Geweke	215
2.2. Метод Gelman и Rubin	216

ПРОЛОГ

ХОЧУ ИЗУЧАТЬ
БАЙЕСОВСКУЮ
СТАТИСТИКУ!



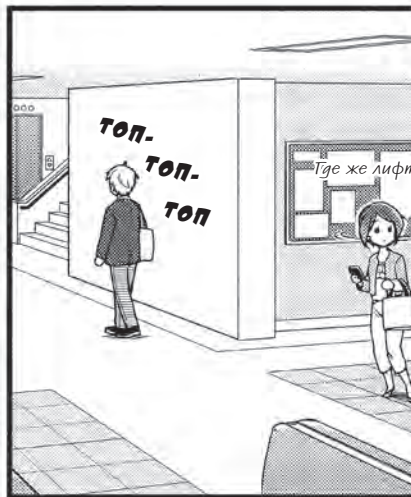
А ВОТ И УНИВЕРСИТЕТ,
ГДЕ РАБОТАЕТ
АКАНЕ-СЕНСЕЙ...

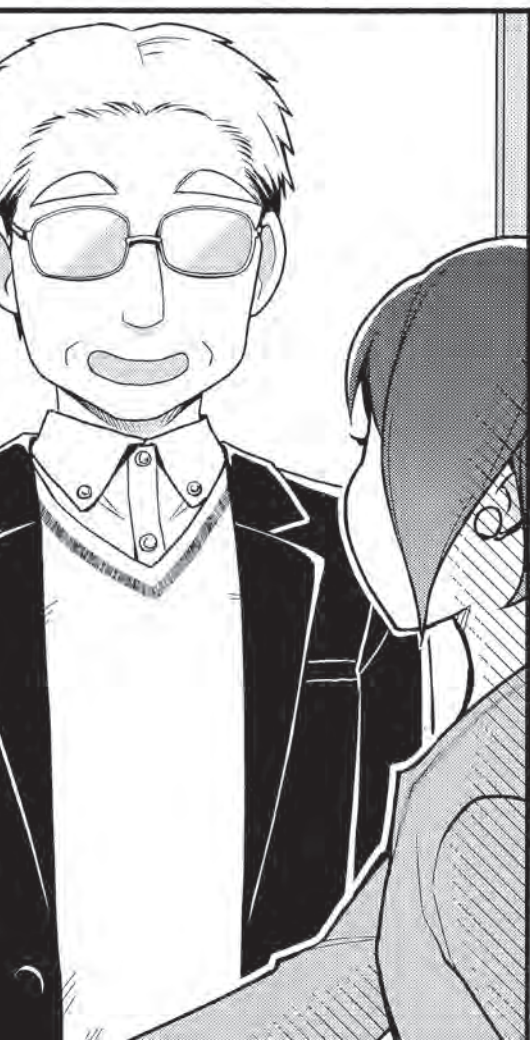
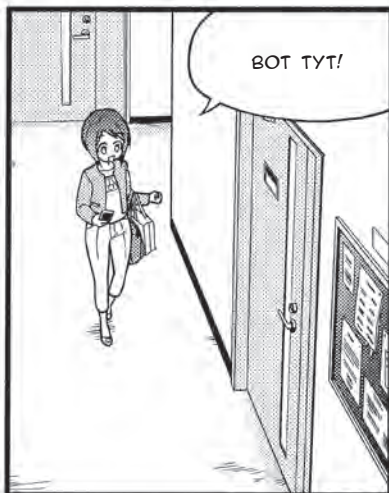
Конно Нанами

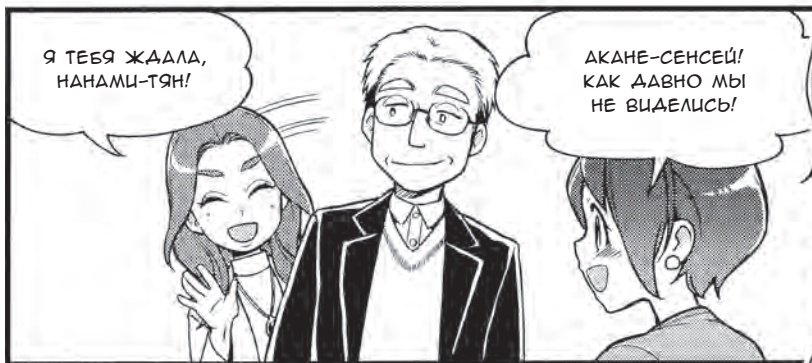
В МЛАДШЕЙ ШКОЛЕ ОНА
БЫЛА МОИМ РЕПЕТИТОРОМ,
А СЕЙЧАС УЖЕ ПРЕПОДАЁТ
В УНИВЕРСИТЕТЕ...

Вот как!

АКАНЕ-СЕНСЕЙ ВСЕГДА
БЫЛА ДОБРОЙ И ТАК ЗАОРОВО
ВСЁ ОБЪЯСНЯЛА.







Я ТЕБЯ ЖАЛАЛ,
НАНАМИ-ТЯН!

АКАНЕ-СЕНСЕЙ!
КАК ДАВНО МЫ
НЕ ВИДЕЛИСЬ!



В ПОСЛЕДНИЙ РАЗ
МЫ ВИДЕЛИСЬ,
КАЖЕТСЯ, ВО ВРЕМЯ
ВСТУПИТЕЛЬНЫХ
ЭКЗАМЕНОВ?

ХОТЯ МЫ И ОБЩАЛИСЬ
ПО ЭЛЕКТРОННОЙ ПОЧТЕ,
НО В ЖИВУЮ Я ТЕБЯ
И ПРАВА ДАВНО
НЕ ВИДЕЛА!

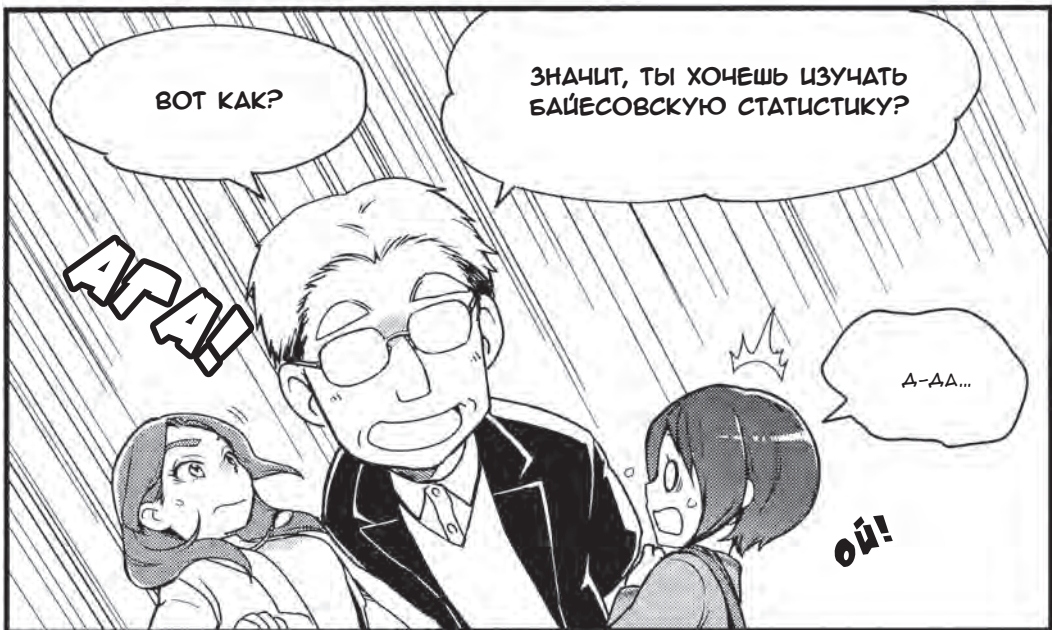
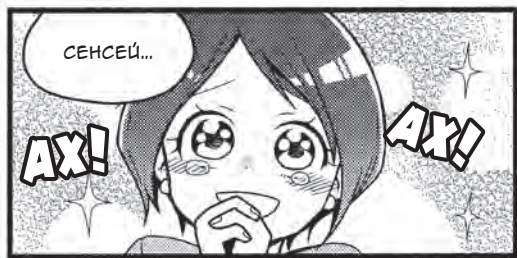
АГА! А ВЫ СОВСЕМ
НЕ ИЗМЕНИЛИСЬ!

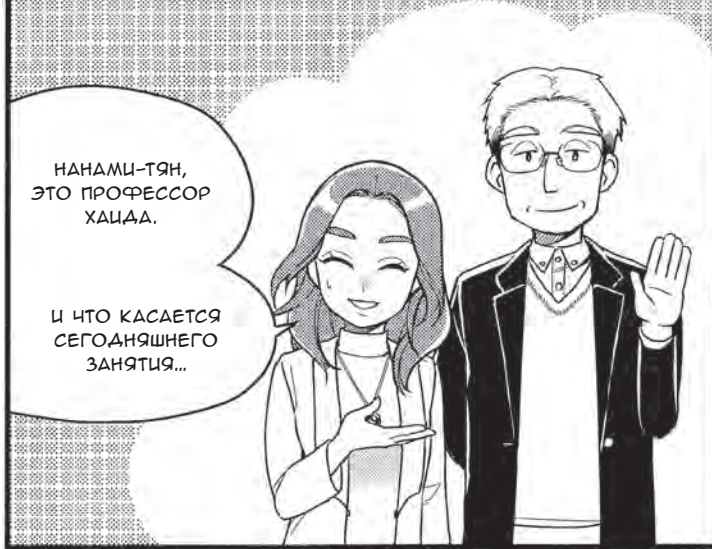


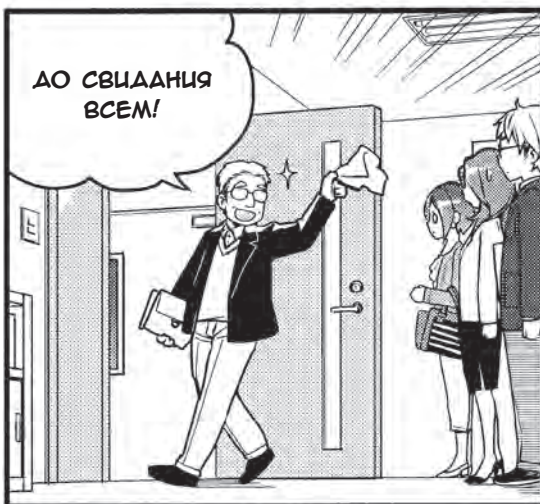
СПАСИБО, ЧТО ВЗЯЛИСЬ
РАССКАЗАТЬ МНЕ
О БАЙЕСОВСКОЙ
СТАТИСТИКЕ!

А ЭТО ВАМ...

Ой,
ЗАЧЕМ ЖЕ ТЫ!
СПАСИБО!







Я, ЧЕСТНО ГОВОРЯ,
ТОЖЕ УАИВИЛАСЬ,
КОГДА ПРОФЕССОР
ХАЦЦА ВАРУГ ПОПРОСИЛ
МЕНЯ ВЗЯТЬ ЕЩЁ
ОДНОГО СТУДЕНТА.

НО МНЕ БЕЗ РАЗНИЦЫ,
УЧИТЬ ОДНОГО ИЛИ
ДВУХ!



ИТАК, РАЗРЕШИТЕ
ЕЩЁ РАЗ ОФИЦИАЛЬНО
ПРЕДСТАВИТЬСЯ.
МЕНЯ ЗОВУТ
НАЦУМИ АКАНЕ.

НАНАМИ-ТЯН,
ЯМАБУКИ-САН,
РАДА ПРИВЕТСТВОВАТЬ
ВАС НА МОЁМ КУРСЕ.



МЕНЯ ЗОВУТ КОННО НАНАМИ,
РАДА ЗНАКОМСТВУ!

ПОКЛОН



МЕНЯ ЗОВУТ
ЯМАБУКИ ООСУКЕ,
Я ЗАНИМАЮСЬ
ИССЛЕДОВАТЕЛЬСКОЙ
РАБОТОЙ.

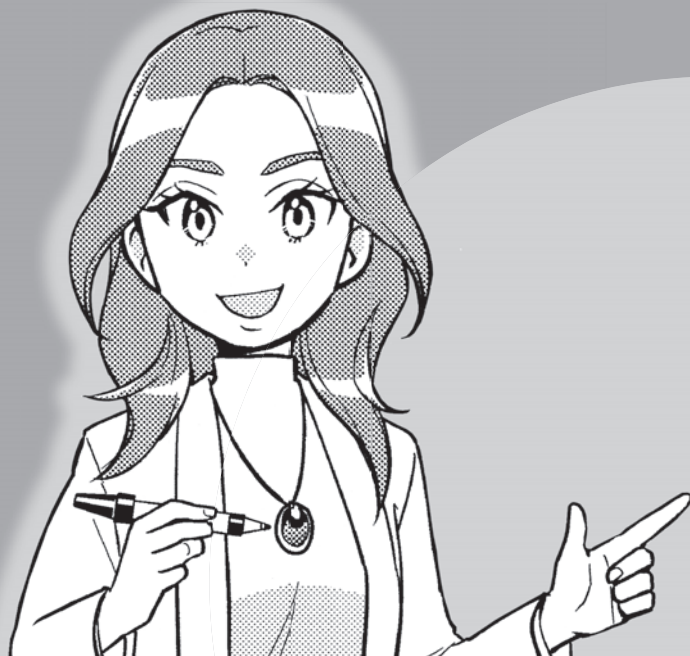
РАДА ЗНАКОМСТВУ.



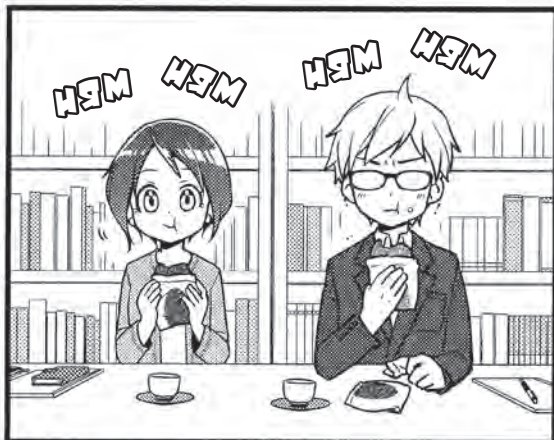
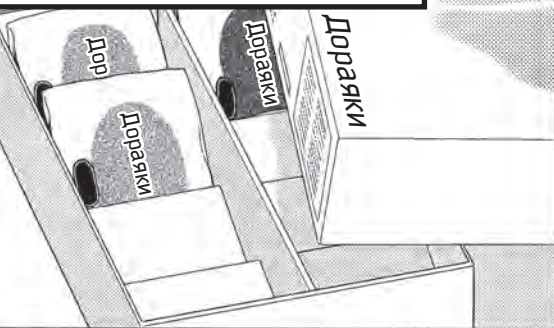


ГЛАВА 1

ЧТО ТАКОЕ БАЙЕСОВСКАЯ СТАТИСТИКА?



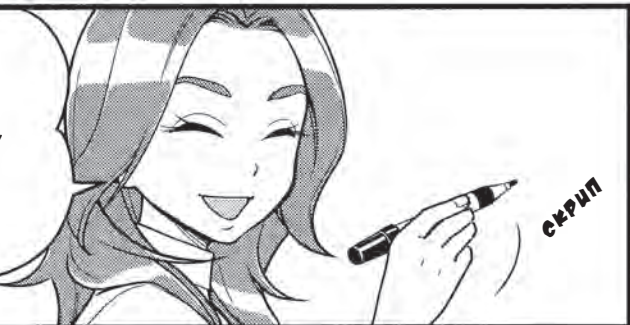
1.1. БАЙЕСОВСКАЯ СТАТИСТИКА



Дораяки – вид японских десертов (два бисквита с начинкой между ними). – Прим. перев.



НЕТ, БАЙЕСОВСКАЯ СТАТИСТИКА НЕ ЯВЛЯЕТСЯ ПРОСТО НАЗВАНИЕМ МЕТОДА АНАЛИЗА. МОЖНО СКАЗАТЬ, ЧТО ТРАДИЦИОННАЯ СТАТИСТИКА И БАЙЕСОВСКАЯ СОСТАВЛЯЮТ СВОЕОБРАЗНУЮ ПАРУ.



ВОТ ПРИМЕРЫ ОТНОШЕНИЙ, ПОХОЖИХ НА ОТНОШЕНИЕ МЕЖДУ ТРАДИЦИОННОЙ И БАЙЕСОВСКОЙ СТАТИСТИКОЙ.

Традиционная статистика

Байесовская статистика

Правая рука
Токио
Удон с водорослями



Левая рука
Осака
Удон с рыбой

...КАК-ТО ТАК.

ЧТО?! УАОН?!



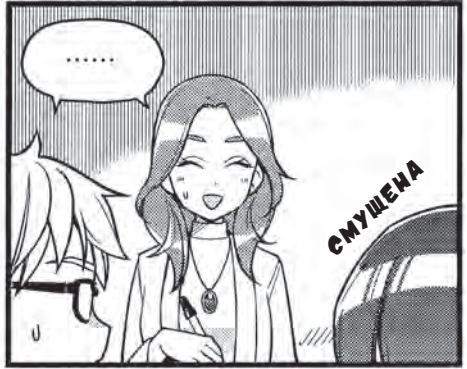
ТО ЕСТЬ НЕЛЬЗЯ СКАЗАТЬ, ЧТО ОДНО ЛУЧШЕ ДРУГОГО, ДА? НЕТ ВОЗМОЖНОСТИ СРАВНИВАТЬ ТАКИЕ ВЕЩИ.

ПРАВИЛЬНО!

И В ДОКАЗАТЕЛЬСТВО ЭТОМУ ХОЧУ СКАЗАТЬ, ЧТО РЕГРЕССИОННЫЙ И ФАКТОРНЫЙ АНАЛИЗЫ ИСПОЛЗУЮТСЯ И НА ОСНОВАНИИ БАЙЕСОВСКОЙ СТАТИСТИКИ!







В ПОДОБНОМ
СЛУЧАЕ,

КОГДА РЕЧЬ ИДЁТ
О ЛЮДЯХ...

- Принимая во внимание привычку опаздывать и годовщину, вероятность того, что молодой человек придёт вовремя, равна примерно $1/2$.
- Поскольку это годовщина, то на этот раз он не опоздает. То есть вероятность прихода вовремя равна 1.



...ВЕРОЯТНОСТЬ ЧАСТО
РАССМАТРИВАЕТСЯ КАК СТЕПЕНЬ
ПЕРСОНАЛЬНОЙ УВЕРЕННОСТИ.

И В САМОМ ДЕЛЕ!
ОБЫЧНО...

Вероятность того, что новое
кафе продержится год, где-то
пятьдесят на пятьдесят.

А когда получишь
очень хорошие результаты
первого вступительного
экзамена, то вероятность того,
что тебя допустят на второй
экзамен, - 99 %!

Точно
получится!

Студенческий
номер: 013
Имя:
Колло. Нанами
94

...КАК-ТО ТАК МЫ
И ОЦЕНИВАЕМ
ВЕРОЯТНОСТЬ.

ВОТ ВИДИШЬ!

ВЕРОЯТНОСТЬ, ОПРЕДЕЛЯЕМАЯ
СТЕПЕНЬЮ УВЕРЕННОСТИ ИНДИВИДА,
НАЗЫВАЕТСЯ СУБЪЕКТИВНОЙ
ВЕРОЯТНОСТЬЮ.



Байесовская статистика

Вероятность = степень персональной уверенности

ЭТО И ЕСТЬ БАЙЕСОВСКАЯ
СТАТИСТИКА.

ИНТЕРЕСНАЯ
КОНЦЕПЦИЯ.

КСТАТИ, БАЙЕС
БЫЛ СВЯЩЕННИКОМ
В XVIII ВЕКЕ.

А КАК СВЯЩЕННИК
ОКАЗАЛСЯ СВЯЗАН
С МАТЕМАТИКОЙ?

ой!

ПРО ЭТО Я И САМА
НЕ ОЧЕНЬ-ТО
ЗНАЮ...

Благославляю
тебя,
Нанамы!

Интересно...



А?! ВОТ КАК...

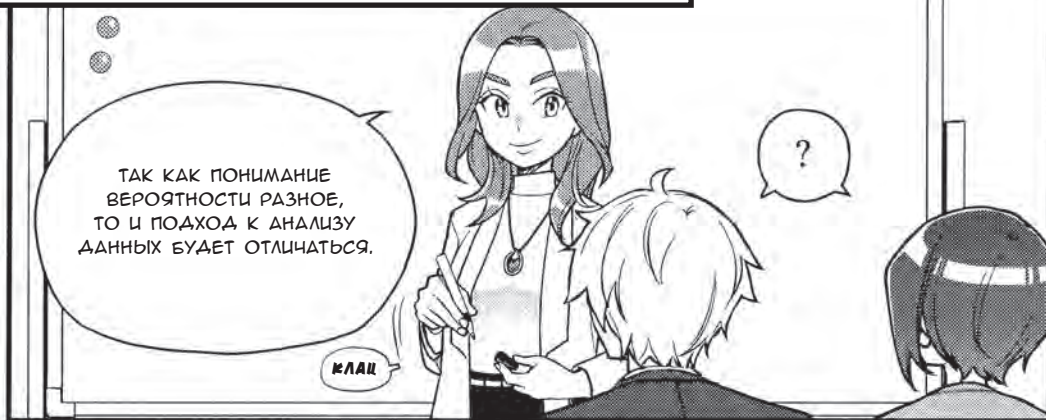
Я ТОГАА ПРО ЭТО ПОТОМ
ПОИЩУ ИНФОРМАЦИЮ!

.....



(Примечание по поводу лица Байеса: так дети в Японии схематически рисуют лица элементами хираганы – японской слоговой азбуки.)

1.2. РАЗЛИЧИЕ МЕЖДУ ОБЫЧНОЙ И БАЙЕСОВСКОЙ СТАТИСТИКАМИ



Общая статистика

1. Из генеральной совокупности случайным образом выбирается n человек, и находится среднее значение для данной группы \bar{x}_1 . Затем возвращаемся к генеральной совокупности.



2. Если действия из пункта 1 проводились T раз, то значение μ будет приблизительно равно $(\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_T)/T$. Но в реальном мире трудно осуществлять действия из пункта 1 многократно, поэтому обычно $T = 1$ и значение μ получают на основании значения $x - 1$.



ЕСЛИ $\bar{x}_1 = 26\,944$,
ТО μ РАВНО...

Байесовская статистика

1. Делается предположение на основании общих знаний и опыта.



μ ТОЧНО БОЛЬШЕ
0 ЦЕН И МЕНЬШЕ
100 ТЫСЯЧ ЦЕН

2. Из генеральной совокупности выбирается n человек (случайным образом), и на основании их данных вычисляется среднее значение \bar{x} .



3. На основании предположения из пункта 1 и значения \bar{x} из пункта 2 применяем так называемую **теорему Байеса**, производим вычисления и получаем результат.



μ БОЛЬШЕ 25 020 ЦЕН
И МЕНЬШЕ 28 940 ЦЕН
С СУБЪЕКТИВНОЙ
ВЕРоятНОСТЬЮ 0,7!

В ПРИВЕДЁННОМ ПРИМЕРЕ
В РЕЗУЛЬТАТЕ ПРИМЕНЕНИЯ
БАЙЕСОВСКОЙ СТАТИСТИКИ
МЫ ПОЛУЧИЛИ ОЦЕНОЧНОЕ
ЗНАЧЕНИЕ μ
И СУБЪЕКТИВНУЮ
ВЕРОЯТНОСТЬ.

Но это
не обязательно
всегда так!

ВКЛЮЧЕНИЕ
В АНАЛИЗ
ПРЕДПОЛОЖЕНИЯ
ИССЛЕДОВАТЕЛЯ
ВЫГЛЯДИТ
НЕСКОЛЬКО
САМОНАДЕЯННО...

СКОРЕЕ, ЕСТЬ ЧУВСТВО
АКТИВНОЙ ПОЗИЦИИ, НЕ ТАК ЛИ?

**ЕСТЬ!
ДАЖЕ ОЧЕНЬ!**

ХИ-ХИ.
НАНАМИ-ТЯН,
ТЫ ВСЁ ТАКАЯ ЖЕ,
КАК В ДЕТСТВЕ!

...А ВЫ, ЗНАЧИТ,
ДАВНО ЗНАКОМЫ?

АКАНЕ-СЕНСЕЙ БЫЛА
МОИМ РЕПЕТИТОРОМ,
КОГДА Я УЧИЛАСЬ ЕЩЁ
В МЛАДШЕЙ ШКОЛЕ!

НАНАМИ-ТЯН
С ДЕТСТВА БЫЛА ТАКОЙ
УВЛЕЧЁННОЙ УЧЕНИЦЕЙ.

ДА ЕЩЁ У НАС ДЕНЬ
РОЖДЕНИЯ В ОДИН ДЕНЬ.
ТАК ЧТО У НАС МНОГО
ОБЩИХ ВОСПОМИНАНИЙ.



ВОТ ОНО ЧТО...



ТАК КАК СЕГОДНЯ
У НАС ПЕРВОЕ ЗАНЯТИЕ,
ТО ДАВАЙТЕ НА ЭТОМ
ЗАКОНЧИМ.

В СЛЕДУЮЩУЮ
СУББОТУ БУДУ
ВАС ЖААТЬ.

ВСЕГО ДОБРОГО!



ЗЫРК



ЗЫРК



О ЧЁМ БЫ
ПОГОВОРИТЬ?

А ПОЧЕМУ ВЫ
РЕШИЛИ ИЗУЧАТЬ
БАЙЕСОВСКУЮ
СТАТИСТИКУ?



...ДУМАЮ,
ОНА ПРИГОДИТСЯ
В МОЕЙ РАБОТЕ...

И У МЕНЯ
ТО ЖЕ САМОЕ!

.....

...В-ВОТ КАК...

Я В УНИВЕРСИТЕТЕ НЕМНОГО
ИЗУЧАЛА СТАТИСТИКУ,
И СЕЙЧАС МНЕ ЭТО
ОКАЗАЛОСЬ НУЖНО
ДЛЯ РАБОТЫ. И НАЗВАНИЕ
"БАЙЕСОВСКАЯ СТАТИСТИКА"
СТАЛО ВСЁ ЧАЩЕ
ПОПАДАТЬСЯ.

Я ПРОБОВАЛА ЧИТАТЬ
КНИГИ НА ЭТУ ТЕМУ,
НО ЧТО-ТО НЕ ОЧЕНЬ
ХОРОШО ИХ ПОНЯЛА...

ПОНИМАЮ...

Я МАБУКИ-САН,
ВЫ ТОЖЕ ИСПЫТЫВАЕТЕ
ТАКИЕ ТРУДНОСТИ, ДА?

НА СЛЕДУЮЩЕЙ НЕДЕЛЕ
БУДЕМ ВМЕСТЕ
СТАРАТЬСЯ!

...ХОРОШО.



