

Дз 1. Симуляции, 5 фактов и бонус

Винни-Пух

7 октября 2013 г.

1 Ворожба и колдовство

Эти шаманские заклинания загружают дополнительные пакеты R: ‘xtable’ — для переделки табличек в латех, ‘ggplot2’ — для красивых графиков, ‘knitr’ — для настройки взаимодействия R — латех. Оставьте это шаманство как есть.

```
require(knitr)
require(xtable)

## Loading required package: xtable

require(ggplot2)

## Loading required package: ggplot2

opts_chunk$set(dev = "png", dpi = 300)
```

2 Обо мне

Меня зовут Винни-Пух.

1. Меня пугают длинные слова
2. У меня в голове — опилки
3. День рождения — 21 августа 1921 года
4. Обожаю играть в Пустяки
5. Открыл Северный полюс



3 Симуляции

Из колоды в 52 карты случайным образом выбирается 13 карт. Случайная величина X — количество выбранных тузов, событие A — выбран туз пик. С помощью компьютерных симуляций оцените $P(X \geq 2)$, $P(X \geq 2 \mid X \geq 1)$, $P(X \geq 2 \mid A)$, $E(X)$, $E(X|A)$.

Решение.

Определим полезные функции.

```
deal <- function(n) {  
  # this function will deal n cards  
  return(sample(1:52, n))  
}  
n.ace <- function(n) {  
  # deal n cards and calculate the number of aces  
  deck <- deal(n)  
  # calculate remainders  
  remainders <- deck%%13  
  # we agreed that aces are the cards with remainder zero  
  is.ace <- remainders == 0
```

```

    return(sum(is.ace))
}
info <- function(n) {
  deck <- deal(n)
  remainders <- deck%%13
  is.ace <- remainders == 0
  is.ace.spade <- deck == 13
  return(c(sum(is.ace), sum(is.ace.spade)))
}

```

Поехали!

```

n.exps <- 10^4
x <- replicate(n.exps, n.ace(13))

```

1. $\hat{P}(X \geq 2)$

```

sum(x >= 2)/n.exps

## [1] 0.2608

```

2. $\hat{P}(X \geq 2 \mid X \geq 1)$

```

sum(x >= 2)/sum(x >= 1)

## [1] 0.3743

```

Для сравнения — настоящая вероятность, $P(X \geq 2 \mid X \geq 1)$

```

p0 <- choose(48, 13)/choose(52, 13) # P(X=0)
p1 <- choose(4, 1) * choose(48, 12)/choose(52, 13) # P(X=1)
(1 - p0 - p1)/(1 - p0) # P(X >= 2)/P(X >= 1)

## [1] 0.3696

```

3. $\hat{E}(X)$

```
mean(x)

## [1] 1.004
```

С событием A чуть посложнее, по каждому опыту нам нужно знать, сколько там было тузов и выпал ли туз пик.

```
xx <- replicate(n.exps, info(13))
```

Транспонируем и превратим в табличку с данными

```
df <- data.frame(t(xx))
names(df) <- c("x", "a")
head(df)

##    x a
## 1 1 1
## 2 0 0
## 3 1 1
## 4 1 0
## 5 0 0
## 6 0 0
```

Решаем остатки:

1. $\hat{E}(X | A)$

```
mean(df$x[df$a == 1])

## [1] 1.717
```

2. $\hat{P}(X \geq 2 | A)$

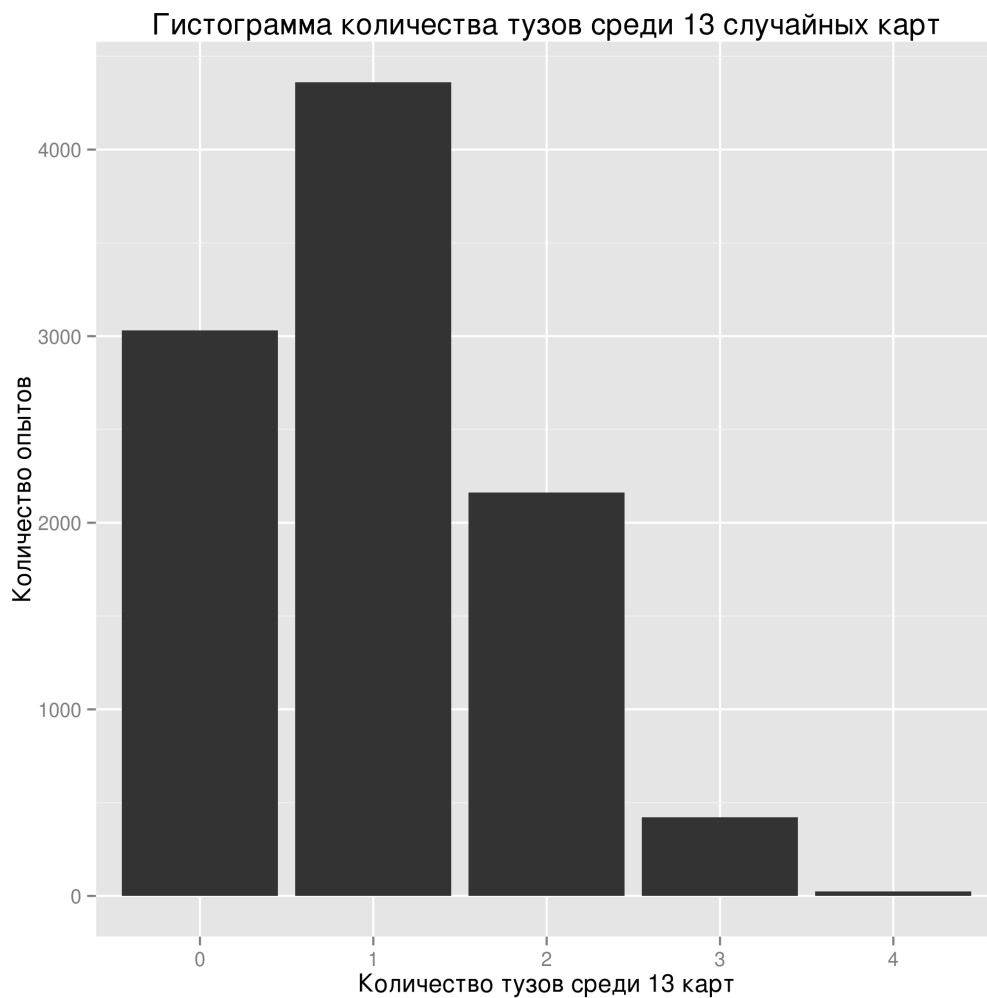
```
sum(df$x >= 2 & df$a == 1)/sum(df$a == 1)

## [1] 0.5713
```

Обратите внимание на крышечку над \hat{P} и \hat{E} . Это не истинные вероятности и математические ожидания, а их оценки.

Бонус. Гистограмма для величины X

```
qplot(as.factor(x), main = "Гистограмма количества тузов среди 13 случайных карт",
      xlab = "Количество тузов среди 13 карт", ylab = "Количество опытов")
```



4 Таблички и графики

Красиво выводим табличку! В R есть куча встроенных наборов данных, например данные по длине тормозного пути и скорости машин 100 лет назад.

```
df <- cars
df.start <- head(df)
xtable(df.start)
```

	speed	dist
1	4.00	2.00
2	4.00	10.00
3	7.00	4.00
4	7.00	22.00
5	8.00	16.00
6	9.00	10.00

Симпатявый график.

```
qplot(speed, dist, data = df, xlab = "Скорость машины, миль в час",
      ylab = "Длина тормозного пути, футы", main = "Зависимость длины тормозного п
```

