

## Задача

Преподаватель хочет провести опрос студентов, с целью узнать процент списывающих. Так как надежда на то, что студенты честно ответят на такой вопрос, не велика, он придумал следующую процедуру для обеспечения анонимности опроса.

При ответе на вопрос студент сначала подбрасывает монетку. Если выпадает орёл, то студент честно отвечает, списывал он или нет. Если выпадает решка, то студент бросает монету второй раз и использует её для ответа: он отвечает «да, списывал» если на монете выпал орёл и «нет, не списывал», если решка (при этом совершенно не важно, списывал он или нет на самом деле).

В таком варианте студенты могут быть спокойны – получив ответ «да» преподаватель не знает, является ли это «да» честным ответом или результатом подбрасывания монетки.

Пусть преподаватель опросил 100 студентов и получил 35 ответов «да». Требуется оценить процент списывающих студентов.

## Модель

Введём следующие обозначения:

$p_{cheat}$  – процент списывающих студентов (частота списывания) – параметр, который нужно найти;

$N$  – число студентов, участвовавших в опросе (100);

*Observation* – полученное наблюдение, т.е. число ответов «да» (35).

В этой задаче можно вычислить частоту  $p_{yes}$  получения ответа «да» в опросе, если известна настоящая частота списывания  $p_{cheat}$ :

$$p_{yes} = P(\text{на первой монете выпал орёл}) * p_{cheat} + P(\text{на первой монете решка}) * P(\text{на второй монете орёл}) = \frac{1}{2} * p_{cheat} + \frac{1}{2} * \frac{1}{2}.$$

Учитывая это, для задачи о списывании можно составить следующую модель:

$$p_{cheat} \sim \text{Uniform}(0,1);$$

$$p_{yes} = \frac{p_{cheat}}{2} + \frac{1}{4};$$

$$\text{Observation} \sim \text{Binomial}(N, p_{yes});$$

$$N = 100.$$

## Программная реализация

Реализуйте модель в программе.

1. Создайте случайную переменную для  $p_{cheat}$  с равномерным априорным распределением на  $[0, 1]$ .
2. Создайте детерминированную переменную, вычисляющую частоту ответа «да»  $p_{yes}$  по частоте списывания  $p_{cheat}$ .
3. Создайте случайную переменную *Observation* с биномиальным распределением, для числа ответа «да», укажите для неё наблюдаемое значение, равное 35.

4. Создайте модель `model`, включающую переменные  $p_{cheat}$ ,  $p_{yes}$ ,  $Observation$ .
5. Запустите алгоритм вероятностного вывода, используя следующий код:  

```
mcmc=pm.MCMC(model)
mcmc.sample(25000, 2500)
```
6. Получите значения из распределения апостериорной вероятности  $p_{cheat}$ .
7. Постройте гистограмму этого распределения.
8. Посмотрите, как будет изменяться эта гистограмма для  $Observation = 25, 50, 75$ . Чему соответствуют такие значения для числа ответов «Да» в данной задаче?