

Главная мораль: байесовский подход — это всего лишь формула условной вероятности.

1. Задача о целебных лягушках :)

У одного вида лягушек самки обладают целебными свойствами. Самцы и самки встречаются равновероятно. Неподалёку видны аж две лягушки данного вида, но издалека неясно кто.

Определите вероятность того, что среди этих лягушек есть хотя бы одна целебная, в каждой из ситуаций:

- а) Самцы квакают, самки — нет, со стороны лягушек слышно кваканье, но не разобрать, одной лягушки или двух.
- б) Самцы и самки квакают по разному, но одинаково часто. Только что послышался отдельный квак одной из лягушек и это квак самца.

2. Яичный бой

Саша и Маша играют в «яичный бой». Перед ними корзина яиц. В начале боя они берут по одному яйцу и бьют их острыми концами. Каждое яйцо в корзине обладает своей «силой», все силы — разные. Более сильное яйцо разбивает более слабое. Внешне яйца не отличимы. Сила яйца не убывает при ударах. Разбитое яйцо выбрасывают, побеждённый берёт новое, а победитель продолжает играть прежним.

Какова вероятность того, что Маша победит в 11-ом раунде, если она уже победила 10 раундов подряд?

3. Классика жанра

Перед нами определение бета-распределения $Beta(\alpha, \beta)$:

$$f(x) \propto \begin{cases} x^{\alpha-1}(1-x)^{\beta-1}, & \text{если } x \in [0; 1] \\ 0, & \text{иначе.} \end{cases}$$

Блондинка Анжелика хочет оценить неизвестную вероятность встретить динозавра, p . Она предполагает, что динозавры встречаются каждый день независимо от других с постоянной вероятностью. Априорно Анжелика считает, что неизвестное p имеет бета-распределение $Beta(2, 3)$. За 20 дней Анжелика 5 раз видела динозавра. Для краткости обозначим вектором y все имеющиеся наблюдения. Величина y_i — результат i -го дня: 1, если динозавр встретился, и 0 иначе.

- а) Чему, по-мнению Анжелики, равны априорные $E(p)$, мода распределения p ?
- б) Найдите апостериорное распределение $f(p|y)$.
- в) Найдите апостериорные ожидание $E(p|y)$ и моду.
- г) Найдите условное распределение y_{21} с учётом имеющихся данных.

4. Рассмотрим следующий код в stan.

```
data {
  int<lower=1> N_x;
  int<lower=1> N_y;
  real y[N_y];
  real x[N_x];
}
parameters {
  real mu_x;
  real mu_y;
  real<lower=0> sigma_x;
  real<lower=0> sigma_y;
}
model {
  for (n_x in 1:N_x) {
    x[n_x] ~ normal(mu_x, sigma_x);
  }
  for (n_y in 1:N_y) {
    y[n_y] ~ normal(mu_y, sigma_y);
  }
  mu_x ~ normal(0, 100);
  mu_y ~ normal(0, 100);
  sigma_x ~ exponential(50);
  sigma_y ~ exponential(50);
}
generated quantities {
  delta = mu_x - mu_y;
  ratio = sigma_x / sigma_y;
}
```

- а) Выпишите предполагаемую модель для данных.
- б) Выпишите априорное распределение.
- в) Байесовский интервал для каких величин позволяет построить данный код?
- г) Какие предпосылки мешают применить в данном случае классический доверительный интервал для разности математических ожиданий, основанный на F -распределении?

5. Просто красивая задача про выборку :)

Есть неизвестное количество чисел. Среди этих чисел одно число встречается строго больше 50% раз. Ведущий показывает числа исследователю Акану в некотором порядке. Когда все числа закончатся, ведущий скажет «всё». Задача Акана — определить, какое число встречается чаще всех. Проблема в том, что Акан так готовился к контрольной по теории вероятностей, что устал. И больше 10 чисел запомнить не в состоянии.

Предложите алгоритм, который позволит Акану определить искомое число.