

Домашнее задание 2. Сопряженные распределения и обоснованность модели.

Курс: Байесовские методы в анализе данных, 2019

Пусть $X = \{x_1, x_2, x_3, \dots, x_N\}$ — N независимых подбрасываний игрального кубика, $x_i \in \{1, 2, 3, 4, 5, 6\}$. Обозначим N_k — число раз, когда выпало число k , $k = 1, \dots, 6$. Тогда правдоподобие модели (мультиномиальное распределение) выглядит следующим образом:

$$p(X|\theta) = \frac{N!}{N_1! \dots N_k!} \prod_{k=1}^6 \theta_k^{N_k},$$

где $\theta_k \geq 0$ — вероятность выпадения числа k , $k = 1, \dots, 6$, $\sum_{k=1}^6 \theta_k = 1$ (*). Требуется:

1. найти оценку максимального правдоподобия θ_{ML} (не забудьте про ограничение (*));
2. подобрать сопряжённое априорное распределение $p(\theta)$ и обосновать, почему вы его таким выбрали (привести рассуждения про функциональную форму распределения). Подсказка: сопряженное распределение будет одним из набора: Парето-распределение, распределение Дирихле, геометрическое распределение (почитайте про все эти распределения в Википедии);
3. найти апостериорное распределение $p(\theta|X)$ (обосновать, почему оно такое);
4. вычислить байесовскую оценку для θ как мат.ожидание $p(\theta|X)$;
5. посчитать обоснованность модели;
6. предложить пример из жизни, где правдоподобие могло бы задаваться тем распределением, которое в этой задаче является априорным (ответ на пункт 2).