Прикладная статистика в машинном обучении

Семинар 13

---

8 декабря 2020 г.

Задача 1.

Пусть $X_1, \ldots, X_n$ – выборка независимых одинаково распределённых случайных величин, имеющих распределение Бернулли с параметром p. Рассмотрим в качестве априорного распределения параметра p равномерное на подходящем интервале, то есть $f(p) = 1$.

a) Выпишите апостериорное распределение p с точностью до константы.

b) Бета-распределение с параметрами α и β задаётся плотностью:

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

Подберите такие параметры α и β, чтобы показать, что полученное апостериорное распределение представимо в виде $p|x \sim C\mathrm{Beta}(\alpha, \beta)$, где C – некоторая константа.

c) Восстановите константу C.

d) Найдите точечную оценку параметра p.

e) Заметим, что $\bar{p}$ представляется в виде

$$\bar{p} = \frac{n}{n+2}\frac{s}{n} + \left(1 - \frac{n}{n+2}\right)\frac{1}{2}.$$

Прокомментируйте интуицию такого разложения.

f) Запишите формулу расчёта 95%-го апостериорного доверительного интервала для p.

g) Теперь предположим, что априорное распределение параметра p – это Бета-распределение $p \sim \mathrm{Beta}(\alpha, \beta)$. Вычислите апостериорное распределение p.

Задача 2.

Рассмотрим выборку независимых одинаково распределённых нормальных случайных величин $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Для простоты будем считать, что σ известна. Пусть априорное распределение параметра μ имеет вид $\mu \sim \mathcal{N}(a, b^2)$.

a) Пусть функция правдоподобия строится лишь по наблюдению X_1. Выведите примерное апостериорное распределение параметра μ. Является ли априорное распределение сопряжённым?

b) Можно показать, что в общем случае апостериорное распределение μ имеет вид

$$\mu|X \sim \mathcal{N}(\bar{\mu}, w^2),$$

где $\bar{\mu} = \lambda\bar{X} + (1-\lambda)a$, $\lambda = \dfrac{1/s^2}{1/s^2 + 1/b^2}$, $s = \sigma/\sqrt{n}$, $1/w^2 = 1/s^2 + 1/b^2$.

Что происходит с λ и w/s при $n \to \infty$? Выпишите приблизительное апостериорное распределение μ при $n \to \infty$.

c) Постройте 95% асимптотический байесовский доверительный интервал для μ.

Задача 3.

Пусть $X_1, \ldots, X_n$ – выборка независимых одинаково распределённых случайных величин из распределения Бернулли с параметром p. Апостериорное распределение задаётся как $f(p) = 1$. Пусть $\psi = \log\left(\frac{p}{1-p}\right)$. Найдите апостериорную функцию распределения и функцию плотности ψ.

Задача 4.

Компания тестирует новый вид удобрений, для чего создаёт контрольную группу размером n_1 и тестовую группу размером n_2. После проведения эксперимента оказалось, что в контрольной группе лучший рост показали X_1 растений, а в тестовой – X_2 растений. Пусть $f(p_1, p_2) = 1$.

a) Выведите апостериорное распределение $f(p_1, p_2|x_1, x_2)$ с точностью до константы.

b) Являются ли $p_1|x_1$ и $p_2|x_2$ независимыми?

c) Опишите процесс байесовской оценки величины $p_2 - p_1$ через симуляцию.