ВМК

Семинар 2. Сопряжённые распределения и экспоненциальный класс распределений

Курс: Байесовские методы в машинном обучении, 2015

Задачи на семинаре:

1. Пусть $x_1, x_2, \dots, x_N$ – независимая выборка из экспоненциального распределения с плотностью $p(x|\lambda) = \lambda \exp(-\lambda x)$, $x \geq 0, \lambda > 0$. Требуется найти оценку максимального правдоподобия λ_{ML}, подобрать сопряжённое распределение $p(\lambda)$, найти апостериорное распределение $p(\lambda|x_1, \dots, x_N)$ и вычислить байесовскую оценку для λ как мат.ожидание $p(\lambda|x_1, \dots, x_N)$.
2. Пусть в N независимых испытаниях Бернулли с вероятностью успеха q произошло k успехов. Требуется подобрать сопряжённое распределение $p(q)$, найти апостериорное распределение $p(q|k, N)$ и его мат.ожидание.
3. На основе бета-распределения подобрать априорное распределение для вероятности q выпадения орла у двухсторонней монетки для следующих ситуаций:
   - «честная» монетка (вероятность q близка к 0.5);
   - «вырожденная» монетка (вероятность q близка к крайним значениям 0 и 1);
   - «нечестная» монетка (вероятность q далека как от 0.5, так и от крайних значений 0 и 1).

   Можно ли аналитически вычислить параметры a, b бета-распределения, если известны его мат.ожидание и дисперсия?
4. Записать плотность гамма-распределения $\mathcal{G}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$ как представителя экспоненциального класса распределений. Найти $\mathbb{E}x$ и $\mathbb{E}\log x$ путём дифференцирования нормировочной константы.
5. Записать биномиальное распределение в форме экспоненциального класса распределений. Найти мат.ожидание и дисперсию распределения путём дифференцирования нормировочной константы.

Задачи для самостоятельной работы:

1. Пусть $x_1, x_2, \dots, x_N$ – независимая выборка из непрерывного равномерного распределения $\mathrm{U}[0, \theta]$. Требуется найти оценку максимального правдоподобия θ_{ML}, подобрать сопряжённое распределение $p(\theta)$, найти апостериорное распределение $p(\theta|x_1, \dots, x_N)$ и вычислить его статистики: мат.ожидание, медиану и моду. *Подсказка: задействовать распределение Парето.*
2. Предположим, что вы приезжаете в новый город и видите автобус с номером 100. Требуется с помощью байесовского подхода оценить общее количество автобусных маршрутов в городе. Какая из статистик апостериорного распределения будет наиболее адекватной? Как изменятся оценки на количество автобусных маршрутов при последующем наблюдении автобусов с номерами 50 и 150? *Подсказка: воспользоваться результатами предыдущей задачи.*
3. Пусть необходимо подобрать априорное распределение для вероятности q выпадения орла у монетки в семействе бета-распределений $\mathrm{Beta}(q|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1}(1-q)^{b-1}$. Априорные предпочтения задаются как $\mathbb{E}q = e$, $p(l < q < u) = 0.95$, где e, l, u – заданные числа. Требуется написать программу для вычисления значений a, b. Какие значения получатся в случае $e = 0.15, l = 0.05, u = 0.3$? При интерпретации найденного априорного распределения как апостериорного при байесовском оценивании q, какому числу подбрасываний монетки оно соответствует?
4. Записать распределение Парето $\mathrm{Pareto}(x|a, b) = \frac{ba^b}{x^{b+1}}[x \geq a]$ при фиксированном a в форме экспоненциального класса распределений. Найти $\mathbb{E}\log x$ путём дифференцирования нормировочной константы.