Московский государственный университет имени М. В. Ломоносова.

Теория 1. Сопряжённые распределения и экспоненциальный класс распределений

Чванкина Дарья
Группа 417
Октябрь 2022

1 Задание 1

$x_1, \dots, x_n \sim U[0, \theta]$

- $\theta_{ML}-?$

  Метод максимального правдоподобия:

$$\prod_{i=1}^{n} \frac{1}{\theta^n} \mathbb{I}(x_i \in [0, \theta]) = \frac{1}{\theta^n} \mathbb{I}(x_{(1)} \geq 0) \mathbb{I}(x_{(n)} \leq \theta) \tag{1}$$

$$\theta_{ML} = \arg\max_{\theta} \frac{1}{\theta^n} \mathbb{I}(x_{(1)} \geq 0) \mathbb{I}(x_{(n)} \leq \theta) \tag{2}$$

  Так как чем больше значение θ, тем меньше правдоподобие, то нужно выбрать минимально возможное значение параметра. Следовательно,

$$\theta_{ML} = x_{(n)} \tag{3}$$

- Сопряженное распределение - ?

  $p(X|\theta) \sim U[0, \theta]$

  $p(\theta)-?$

  Пусть априорное распределение параметра θ равное распределению Парето с параметрами k и θ_m.

$$p(\theta) \sim Pareto(k, \theta_m) = \frac{k\theta_m^k}{\theta^{k+1}} \mathbb{I}(\theta \geq \theta_m) \tag{4}$$

  Тогда апостериорное распределение:

$$p(\theta, X) = \frac{p(X|\theta)p(\theta)}{\int p(X|\theta)p(\theta)d\theta} \tag{5}$$

$$p(X|\theta)p(\theta) = \frac{1}{\theta^n} \mathbb{I}(x_{(1)} \geq 0) \mathbb{I}(x_{(n)} \leq \theta) \frac{k\theta_m^k}{\theta^{k+1}} \mathbb{I}(\theta \geq \theta_m) = \frac{k\theta_m^k}{\theta^{(k+n)+1}} \mathbb{I}(x_{(1)} \geq 0) \mathbb{I}(x_{(n)} \leq \theta) \mathbb{I}(\theta \geq \theta_m) \tag{6}$$

$$\int p(X|\theta)p(\theta)d\theta = \int \frac{k\theta_m^k}{\theta^{(k+n)+1}} \mathbb{I}(x_{(1)} \geq 0) \mathbb{I}(x_{(n)} \leq \theta) \mathbb{I}(\theta \geq \theta_m) d\theta =$$
$$= \int_{max(\theta_m, x_{(n)})} \frac{k\theta_m^k \mathbb{I}(x_{(1)} \geq 0)}{\theta^{(k+n)+1}} d\theta = k\theta_m^k \frac{1}{\theta^{n+k}} \frac{1}{-(n+k)} \Bigg|_{max(\theta_m, x_{(n)})}^{+\infty} = \frac{k\theta_m^k}{n+k} \frac{\mathbb{I}(x_{(1)} \geq 0)}{(max(\theta_m, x_{(n)}))^{n+k}} \tag{7}$$

$$
\begin{aligned}
p(\theta|X) = \frac{\frac{k\theta_m^k}{\theta^{(k+n)+1}}\mathbb{I}(x_{(1)} \geq 0)\mathbb{I}(x_{(n)} \leq \theta)\mathbb{I}(\theta \geq \theta_m)}{\frac{k\theta_m^k}{n+k}\frac{\mathbb{I}(x_{(1)}\geq 0)}{(max(\theta_m, x_{(n)}))^{n+k}}} = \\
= \frac{(n+k)(max(\theta_m, x_{(n)}))^{n+k}\mathbb{I}(\theta \geq max(\theta_m, x_{(n)}))}{\theta^{(k+n)+1}} = \\
= Pareto(n+k, max(\theta_m, x_{(n)}))
\end{aligned} \tag{8}
$$

- $p(\theta|X) = Pareto(n+k, max(\theta_m, x_{(n)}))$

  $\mathbb{E}, med, mod-?$

$$
\begin{aligned}
\mathbb{E}p(\theta|X) = \int_{max(\theta_m, x_{(n)})}^{+\infty} \frac{n+k}{\theta^{n+k+1}}(max(\dots)^{n+k}\theta d\theta = \frac{(n+k)(max())^{n+k}}{-(n+k-1)\theta^{n+k-1}}\Bigg|_{max(\dots)}^{+\infty} = \\
= \frac{n+k}{n+k-1}max(\theta_m, x_{(n)})
\end{aligned} \tag{9}
$$

Пусть теперь у нас есть распределение Парето в общем виде: $Pareto(k, x_m)$. Найдем медиану этого распределения:

$$
\begin{aligned}
F(x) = 1 - (\frac{x_m}{x})^k = 0.5 \\
(\frac{x_m}{x})^k = 0.5 \\
\frac{x_m}{x} = (0.5)^{1/k} \\
x = x_m\sqrt[k]{2}
\end{aligned} \tag{10}
$$

Откуда: $\mathrm{med}\, p(\theta|X) = max(\theta_m, x_{(n)})\sqrt[n+k]{2}$.

Найдем моду распределения Парето в общем виде:

$$
\text{mod } p(x) = \max_x p(x) = \max_x \frac{k(x_m)^k}{x^{k+1}}\mathbb{I}(x \geq x_m) \tag{11}
$$

Так как чем больше x, тем меньше значение дроби, то максимум такого выражения будет достигаться при минимально возможном значении x. Значит, $p(x) = x_m$.

Откуда: $\text{mod } p(\theta|X) = max(\theta_m, x_{(n)})$.

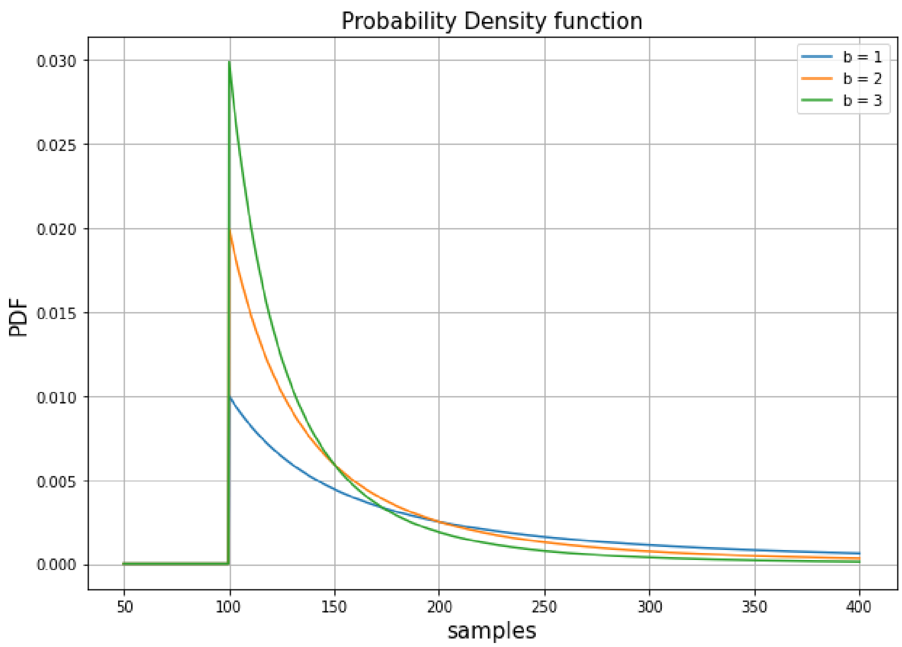
2 Задание 2

Введем случайную величину x - номер автобуса, который мы увидели. Предположим, что пронаблюдать любой номер автобус равновероятно. Тогда x имеет равномерное распределение $U[0, \theta]$, где θ - натуральное число, означающее максимальное количество маршрутов в городе, которое мы хотим оценить. При этом будем считать, что если случайная величина приняла значение в интервале (k, k + 1], то значение маршрута которое мы пронаблюдали равно k + 1. Всего таких отрезков - θ, значит вероятность пронаблюдать конкретное значение номера маршрута равна $\frac{1}{\theta}$.

Пусть априорное распределение параметра θ - распределение Парето с параметрами a и b - $Pareto(a, b) = \frac{ba^b}{x^{b+1}}[x \geq a]$, так как оно является сопряженным к равномерному. Заметим, что в плотности распределения есть индикатор - $[x \geq a]$. Так как мы уже увидели автобус с номером 100, то мы знаем, что маршрутов в городе не меньше 100, следовательно, параметр a следует выбрать не меньше 100. Так как в апостериорном распределении в качестве параметра a выступает максимум из параметра a (из априорного распределения) и максимума наблюдаемых величин, то данный параметр может только увеличиваться при поступлении новых наблюдений. Отсюда логичным положить a равным 100.

Посмотрим на график плотности, как ведет себя распределение Парето при фиксированном параметре a:

Рис. 1:



Так как у нас мало информации о том, сколько может быть маршрутов, насколько большой город, то будет разумным положить b равным 1. Тогда вероятность значения параметра θ быть большим 100 становится больше. При этом из апостериорного распределения (8) следует, что при поступлении новых данных, данный параметр будет только увеличиваться. С увеличением параметра b как видно из графика, мы становимся более уверены в своем прогнозе(график становится более высоким в значении равным a).

Рассмотрим статистики апостериорного распределения. Заметим, что мода распределения всегда равна наибольшему наблюдаемому номеру автобуса. Что не совсем подходит для нашей оценки. Желательно, чтобы при первых наблюдениях оценка колебалась возле максимального значения, то есть была чуть больше. Но при этом при наблюдении 1000 автобусов и при параметре a равным 100 после этих наблюдений, хочется оценивать параметр θ почти равным 100. Как раз этим требованиям удовлетворяет мат. ожидание. Так как параметр θ должен быть натуральным, то для оценки предлагается округлить мат. ожидание до ближайшего целого. Итак, если мы наблюдали всего 5 автобусов (предполагаем, что параметр a

остался равным 100), то мат. ожидание будет равным 125. Если 1000 автобусов, то - 100.

Рассмотрим медиану апостериорного распределения. Из (10) следует, что медианой апостериорного распределения через 1000 наблюдений будет равна(предполагаем, что параметр a остался равным 100) $2^{1/1000}100 \approx 100$. Нас это устраивает, но уже при наблюдении 5 автобусов медиана будет равной ≈ 115, при 10 - ≈ 107. Если сравнивать с мат. ожиданием, то данная функция слишком быстро уменьшает свои значения.

Подтвердим предположения на еще одном примере. Пусть мы пронаблюдали еще два автобуса с номерами 50 и 150. Тогда мат. ожидание становится равным $\frac{3}{2}150 = 225$, медиана - $2^{1/3}150 \approx 189$, мода - 150. Здесь мода - просто наибольшее значение(тогда непонятно зачем применять байесовский вывод), медиана выдает слишком большее значение. Таким образом, в данной задаче наиболее подходящей статистикой является мат. ожидание.

3 Задание 3

Представим распределение Парето с параметрами a и b $\frac{ba^b}{x^{b+1}}[x \geq a]$ при фиксированном a в виде экспоненциального семейства: $\frac{f(x)}{g(\theta)}exp(\theta^T u(x))$.

$$Pareto(a, b) = \frac{ba^b}{x^{b+1}}[x \geq a] = ba^b \exp(-(b+1)\log x)[x \geq a] \tag{12}$$

Тогда

$$\begin{aligned} u(x) &= \log x \\ \theta &= -(b+1) \\ f(x) &= [x \geq a] \\ g(b) &= ((-(-(b+1)) - 1)^{-1} a^{-(-(b+1))-1} \\ g(\theta) &= (-\theta - 1)^{-1} a^{-\theta-1} \end{aligned} \tag{13}$$

$$\frac{\partial \log g(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}(-\log(-\theta-1) + (\theta+1)\log a) = -\frac{1}{\theta+1} + \log a \tag{14}$$

$$\begin{aligned} \mathbb{E}\log x = \frac{\partial \log g(\theta)}{\partial \theta} = -\frac{1}{\theta+1} + \log a = -\frac{1}{-b-1+1} + \log a &= \frac{1}{b} + \log a \\ \mathbb{E}\log x &= \frac{1}{b} + \log a \end{aligned} \tag{15}$$