Московский Государственный
Университет

Сопряжённые распределения и экспоненциальный класс распределений

Выполнил: Курцев Д.В.

Группа: 417

Факультет Вычислительной математики и кибернетики

Кафедра Математических методов прогнозирования

Октябрь 2022

Задача 1

$x_1, ..., x_n \sim U[0,\theta]$. Тогда $p(x) = \frac{1}{\theta}\mathbb{1}\{x \in [0,\theta]\} \quad \Rightarrow \quad$ правдоподобие выборки равно

$$p(X|\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{1}\{x_i \in [0,\theta]\} = \frac{1}{\theta^n}\mathbb{1}\{0 \leq x_{(1)} \leq x_{(n)} \leq \theta\}$$

где за $x_{(i)}$ была обозначена $i-$ая порядковая статистика.
Так как $\frac{1}{\theta^n}$ монотонно убывающая функция, то своё наибольшее значение она принимает в левой границе области определения, т.е. в точке $x_{(n)}$. Таким образом, оценкой максимального правдоподобия является $\theta_{ML} = x_{(n)}$

Найдём сопряжённое распределение $p(\theta)$. Из вида равномерного, можно предположить, что $p(\theta) \sim \theta^k \mathbb{1}(\theta)$. Похожий вид имеет распределение Парето:
$\text{Pareto}(x|a,b) = \frac{ba^b}{x^{b+1}}\mathbb{1}\{\theta \geq a\}$
Найдём апостериорное распредение $p(\theta|X)$.

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{\int p(X|\theta)p(\theta)d\theta} = \frac{1}{Z}p(X|\theta)p(\theta) = \frac{1}{Z}\frac{1}{\theta^n}\mathbb{1}\{\theta \geq x_{(n)}\}\frac{ba^b}{\theta^{b+1}}\mathbb{1}\{\theta \geq a\} =$$

Обозначим за $m = \max\{a, x_{(n)}\}$. Тогда:

$$= \frac{1}{Z}\frac{ba^b}{\theta^{n+b+1}}\mathbb{1}\{\theta \geq m\} = \frac{1}{Z}\frac{ba^b}{(b+n)m^{b+n}}\mathbb{1}\{\theta \geq m\}\ \frac{(b+n)m^{b+n}}{\theta^{n+b+1}} \sim Pareto(m, b{+}n)$$

Таким образом мы доказали, что сопряжённое к равномерному распределению является распределение Парето. Тогда:
$p(\theta) \sim Pareto(a,b), \quad p(\theta|X) \sim Pareto(\max\{a, x_{(n)}\}, b+n)$

Пусть случайная величина $\xi \sim Pareto(a,b)$. Найдём её мат ожидание, медиану и моду.

$$\mathbb{E}\,\xi = \int\limits_a^{+\infty} \frac{ba^b}{x^{b+1}}xdx = ba^b\int\limits_a^{+\infty}\frac{dx}{x^b} = \begin{cases} -\frac{ba^b}{b-1}\frac{1}{x^{b-1}}\big|_a^{+\infty}, & b \neq 1 \\ -ba^b\, lnx\big|_a^{+\infty}, & b = 1 \end{cases}$$

Данный интеграл сходится для $b > 1$. Для таких b получим:

$$\mathbb{E}\,\xi = -\frac{ba^b}{b-1}\frac{1}{x^{b-1}}\bigg|_a^{+\infty} = -0 + \frac{ba^b}{b-1}\frac{1}{a^{b-1}} = \frac{ba}{b-1}$$

Так как распределение Парето является абсолютно непрерывным, то медиана является решением уравнения $\int\limits_a^{med} \frac{ba^b}{x^{b+1}}dx = \frac{1}{2}$

$$\int\limits_a^{med} \frac{ba^b}{x^{b+1}}dx = -\frac{ba^b}{bx^b}\bigg|_a^{med} = \frac{a^b}{a^b} - \frac{a^b}{med^b} = 1 - \frac{a^b}{med^b} = \frac{1}{2} \Rightarrow med = a\sqrt[b]{2}$$

Так как плотность распределения Парето является монотонно убывающей функцие, то своё наибольшее значение она принимает в левой границе области определения, т.е. в точке a. Таким образом, $mod = a$

Задача 2

Можно предположить, что появление автобуса с любого маршрута равновероятно. Поэтому можно считать, что они имеют равномерное распределение на отрезке $[0, \theta]$ $(U[0, \theta])$, где θ - натуральное число. Маршруты начинают нумеровать с 1. Тогда вероятность появления автобуса с номером k равна $\mathbb{P}(x = k) = \int\limits_{k-1}^{k} \frac{1}{\theta}dx = \frac{1}{\theta}(k - (k-1)) = \frac{1}{\theta}$. Аналогично будем считать, для апостериорного распределения: $\mathbb{P}(\theta = k|x) = \int\limits_{k-1}^{k} p(\theta|x)d\theta$.

Тогда, как было получено в прошлой задаче, сопряжённым априорным распределением является распределение Парето. То есть $p(\theta) \sim Pareto(a, b)$, $\;p(\theta|X) \sim Pareto(\max\{a, x_{(n)}\}, b + n)$.

Так как мы находимся на остановке, то следовательно в городе существует хотя бы один маршрут. При этом мы не знаем верхней границы на их количество (вдруг это очень маленький городок с одним автобусом). Поэтому в качестве a можно взять число из $(0, 1]$. Пусть $a = 1$. Большой уверенности, в том, что маршрут всего 1 нет, поэтому b также не особо велико. Пусть $b = 1$.

Тогда после того, как на остановку приходит автобус с номером 100, наша выборка равна $X = \{100\}$. То есть $p_1(\theta|X) \sim Pareto(100, 2)$. А после прихода 50 и 150: $X = \{100, 50, 150\} \;\Rightarrow\; p_2(\theta|X) \sim Pareto(150, 4)$.

В качестве оценки на парметр θ можно взять медиану, как более стабильную статистику. Данную статистику будем округлять вверх. Если предположить, что существует атобус с номером 2.5, то очевидно, что в городе будет более 2 маршрутов, то есть 3 и более. Тогда в первом случае $\theta \approx 141,42 \Rightarrow \theta = 142$, а во втором $\theta \approx 178,38 \quad \Rightarrow \quad \theta = 179.$

Задача 3

Пусть $x \sim Pareto(a, b)$. Тогда:

$$p(x) = \frac{ba^b}{x^{b+1}} \mathbb{1}\{x \geq a\} = ba^b e^{-(b+1)\ln x} \, \mathbb{1}\{x \geq a\}$$

Обозначим тогда $\theta = b + 1, u(x) = -\ln x \quad \Rightarrow \quad g(\theta) = \frac{1}{(\theta-1)a^{\theta-1}}$

$$\mathbb{E} \, \ln x = -\mathbb{E} \, u(x) = -\frac{d}{d\theta} \ln g(\theta) = \frac{d}{d\theta}(\ln(\theta - 1) + (\theta - 1)\ln a) = \frac{1}{\theta - 1} + \ln a =$$

$$\frac{1}{b} + \ln a$$

То есть получим, что $\mathbb{E} \, \ln x = \frac{1}{b} + \ln a$