# Домашнее задание 1 по БММО
# «Сопряжённые распределения и экспоненциальный класс распределений»

***Аят Оспанов***
617 гр., ММП, ВМК МГУ, Москва
22 сентября 2017 г.

## 1 Задача 1

Функция правдоподобия будет иметь следующий вид:

$$p(X|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta}[x_i \leq \theta] = \frac{1}{\theta^n}[\max_{i=1,\ldots,n}(x_i) \leq \theta]$$

Т.к. функция правдоподобия убывает относительно $\theta$, то оценка максимального правдоподобия

$$\theta_{ML} = \max_{i=1,\ldots,n}(x_i)$$

Как можно увидеть из вида функции правдоподобия, сопряженным распределением $p(\theta)$ будет распределение Парето:

$$p(\theta) = Pareto(\theta|a,b) = \frac{ba^b}{\theta^{b+1}}[\theta \geq a]$$

Теперь найдем апостериорное распределение:

$$p(\theta|X) \propto p(X|\theta)p(\theta) = \frac{1}{\theta^n}[\max_{i=1,\ldots,n}(x_i) \leq \theta] \cdot \frac{ba^b}{\theta^{b+1}}[\theta \geq a] =$$

$$\frac{ba^b}{\theta^{n+b+1}}[\max_{i=1,\ldots,n}(a, x_i) \leq \theta] = Pareto(\theta|\max(a, x_1, \ldots, x_n), n+b) = Pareto(\theta|a', b')$$

$$a' = \max(a, x_1, \ldots, x_n)$$

$$b' = n + b$$

Теперь найдем его статистики:

$$\mathbb{E}[\theta|X] = \int\limits_{-\infty}^{+\infty} \theta Pareto(\theta|a', b')d\theta = \int\limits_{-\infty}^{+\infty} \theta \frac{b'a'^{b'}}{\theta^{b'+1}}[\theta \geq a']d\theta = b'a'^{b'} \int\limits_{a'}^{+\infty} \frac{d\theta}{\theta^{b'}} = b'a'^{b'}\big(0 - \frac{a'^{1-b'}}{1-b'}\big) =$$

$$\frac{b'a'}{b'-1} = \frac{(n+b)\max(a, x_1, \ldots, x_n)}{n+b-1}$$

$$P(\theta < \theta_{med}) = \int\limits_{-\infty}^{\theta} \frac{b' a'^{b'}}{\theta^{b'+1}} d\theta = 0.5$$

$$b' a'^{b'} \int\limits_{-\infty}^{\theta} \frac{d\theta}{\theta^{b'+1}} = 0.5$$

$$\left(\frac{a'}{\theta}\right)^{b'} = \frac{1}{2}$$

$$\theta = a' \sqrt[b']{2}$$

$$\theta_{med} = \sqrt[n+b]{2} \max(a, x_1, \dots, x_n)$$

Аналогично с функцией правдоподобия:

$$mode[\theta|X] = arg \max_{\theta} Pareto(\theta|a', b') = a' = \max(a, x_1, \dots, x_n)$$

# 2 Задача 2

Пусть автобусы пронумерованы от 1 до $\theta$. Выборка автобусов берется из равномерного распределния. Тогда априорное распределение $p(\theta)$ есть распределение Парето: $p(\theta) = Pareto(\theta|1, b)$. Тут мы предполагаем, что в городе хотя бы 1 автобус: $\theta \geq 1$.

После первого увиденного автобуса с номером 100, апостериорное распределение $p(\theta|100)$ будет распределение Парето с параметрами $a' = 100$ и $b' = b+1$ и следующей статистикой:

$$\mathbb{E}[\theta|x_1] = \frac{100 * (b+1)}{b} = 100 + \frac{100}{b} \in (100, 200]$$
$$mode[\theta|x_1] = 100$$
$$\theta_{med} = 100 \sqrt[b+1]{2} \in (100, 100\sqrt{2} \approx 142]$$

Мода в данном случае не адекватна, т.к. считать, что автобус с номером 100 будет часто встречающимся не корректно. Матожидание адекватно при некоторых $b$, т.к. если $b = 1$, то считать, что в городе 400 автобусов (2·медиана) после увиденного 100-го не адекватно. Таким образом, в данном случае мода адекватна, т.к. количество автобусов окажется в интервале $(200, 283]$

Теперь рассмотрим случай, когда мы увидели автобусы с номерами 50 и 100. Тогда статистика будет следующей:

$$\mathbb{E}[\theta|x_1, x_2, x_3] = \frac{150 * (b+3)}{b+2} = 150 + \frac{150}{b+2} \in (150, 200]$$
$$mode[\theta|x_1] = 150$$
$$\theta_{med} = 150 \sqrt[b+1]{2} \in (150, 150\sqrt{2} \approx 213]$$

В данном случае мода все еще не адекватна, т.к. мы не видели повторящиеся автобусы. Медиана и матожидание примерно одинаково адекватны, но более адекватно было бы, если бы они были равны 100.

## 3 Задача 3

Запишем распределение Парето при фиксированном $a$ в форме экспоненциального класса:

$$exp(x|b) = \frac{f(x)}{g(b)} e^{bu(x)}$$

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

В итоге получаем, что $f(x) = \frac{[x \geq a]}{x}$, $g(b) = \frac{1}{ba^b}$, $u(x) = -\log(x)$

Найдем $\mathbb{E}\log(x)$:

$$\mathbb{E}\log(x) = -\mathbb{E}u(x) = -\frac{d\log g(b)}{db} = \frac{d\log(ba^b)}{db} = \frac{d(\log(b) + b\log(a))}{db} = \frac{1}{b} + \log(a)$$