Теория 1. Сопряжённые распределения и экспоненциальный класс распределений

Курс: Байесовские методы в машинном обучении, 2020

1. Пусть $x_1, x_2, \ldots, x_N$ – 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 $p(\theta|x_1, \ldots, x_N)$ и вычислить его статистики: мат.ожидание, медиану и моду. Формулы для статистик нужно вывести, а не взять готовые. *Подсказка: задействовать распределение Парето.*

Оценка максимального правдоподобия.
Так как элементы выборки незавимы по условию, то правдоподобие можно записать как $L(\theta) = \prod_{i=1}^{n} p(x_i|\theta)$
Отсюда, учитывая, что распределение равномерное, максимальное правдоподобие записывается как

$$\theta_{ML} = \operatorname*{argmax}_x \sum_{i=1}^{n} \log(p(x_i|\theta)) = \operatorname*{argmax}_x(-n \log \theta)$$

Так как вероятность пронаблюдать значение вне отрезка $[0, \theta]$ равна нулю, значит θ как минимум равна максимальному значению в выборке.
Так как $f(\theta) = -n \log \theta$ возрастающая при любых n натуральных, то максимум будет достигаться в минимальновозможном значение θ. То есть в максимальном значении выборки. То есть $\theta_{ML} = \max x_i$
Сопряженное распределение.
Рассмотрим распределение Парето: $p(\theta|\alpha, \beta) = \dfrac{\alpha\beta^{\alpha}}{\theta^{\alpha+1}}[\beta \leq \theta]$. Проверим, будет ли апостериорное распределение иметь такой же вид.
$p(\theta|X) = \dfrac{p(X|\theta)p(\theta|\alpha,\beta)}{\int_0^{+\infty} p(X|\theta)p(\theta|\alpha,\beta)d\theta} = \dfrac{\frac{\alpha\beta^{\alpha}}{\theta^{\alpha+1}}[\beta \leq \theta]\frac{1}{\theta}[0 \leq \max x_i \leq \theta)]}{\int_0^{+\infty} \frac{\alpha\beta^{\alpha}}{\theta^{\alpha+1}}[\beta \leq \theta]\frac{1}{\theta}[0 \leq \max x_i \leq \theta]d\theta} = /m = \max(\beta, \max x_i)/ =$
$= (n+\alpha)m^{n+\alpha}\theta^{-n-\alpha-1}[m \leq \theta] = /\hat{\alpha} = \alpha + n, \hat{\beta} = m/ = \dfrac{\hat{\alpha}\hat{\beta}^{\hat{\alpha}}}{\theta^{\hat{\alpha}+1}}[\hat{\beta} \leq \theta]$, чтд.

Статистики.

Математическое ожидание
$\mathbb{E}\theta = \displaystyle\int_0^{+\infty} \theta p(\theta|X)d\theta = (n+\alpha)m^{n+\alpha}\int_m^{+\infty} \theta^{-n-\alpha}d\theta = /\hat{\alpha} = n+\alpha; \hat{\beta} = m/ = \hat{\alpha}m^{\hat{\alpha}}\int_m^{+\infty}\theta^{-\hat{\alpha}}d\theta = \hat{\alpha}m^{\hat{\alpha}}\frac{1}{\hat{\alpha}-1}m^{-\hat{\alpha}+1} = \frac{\hat{\alpha}m}{\hat{\alpha}-1} = \left(1 + \frac{1}{\hat{\alpha}}\right)m$

Медиана.
Для непрервыных распределений медиана – решение уравнения $F(x) = 0.5$.
Для нашего распределения $F(x) = (n+\alpha)m^{n+\alpha}\int_x^{+\infty}\theta^{-n-\alpha-1}d\theta = \left(\frac{m}{x}\right)^{\hat{\alpha}}$
Обозначим медиану за med, тогда, решив уравнение $\left(\frac{m}{med}\right)^{\hat{\alpha}} = 0.5$, получим, что $med = m \cdot 2^{\frac{1}{\hat{\alpha}}}$

Мода
Так как мода – значение во множестве наблюдений, которое встречается наиболее часто, то можно записать требуемое следующим образом:

$$\operatorname*{argmax}_\theta p(\theta|X) = \operatorname*{argmax}_\theta \theta^{-n-\alpha-1}[m \leq \theta]$$

$\theta^{-n-\alpha-1}[m \leq \theta]$ – невозрастающая функция от θ, так что максимум достигается в точке минимального возможного значения θ, то есть мода равна $m = \max(\beta, \max x_i)$

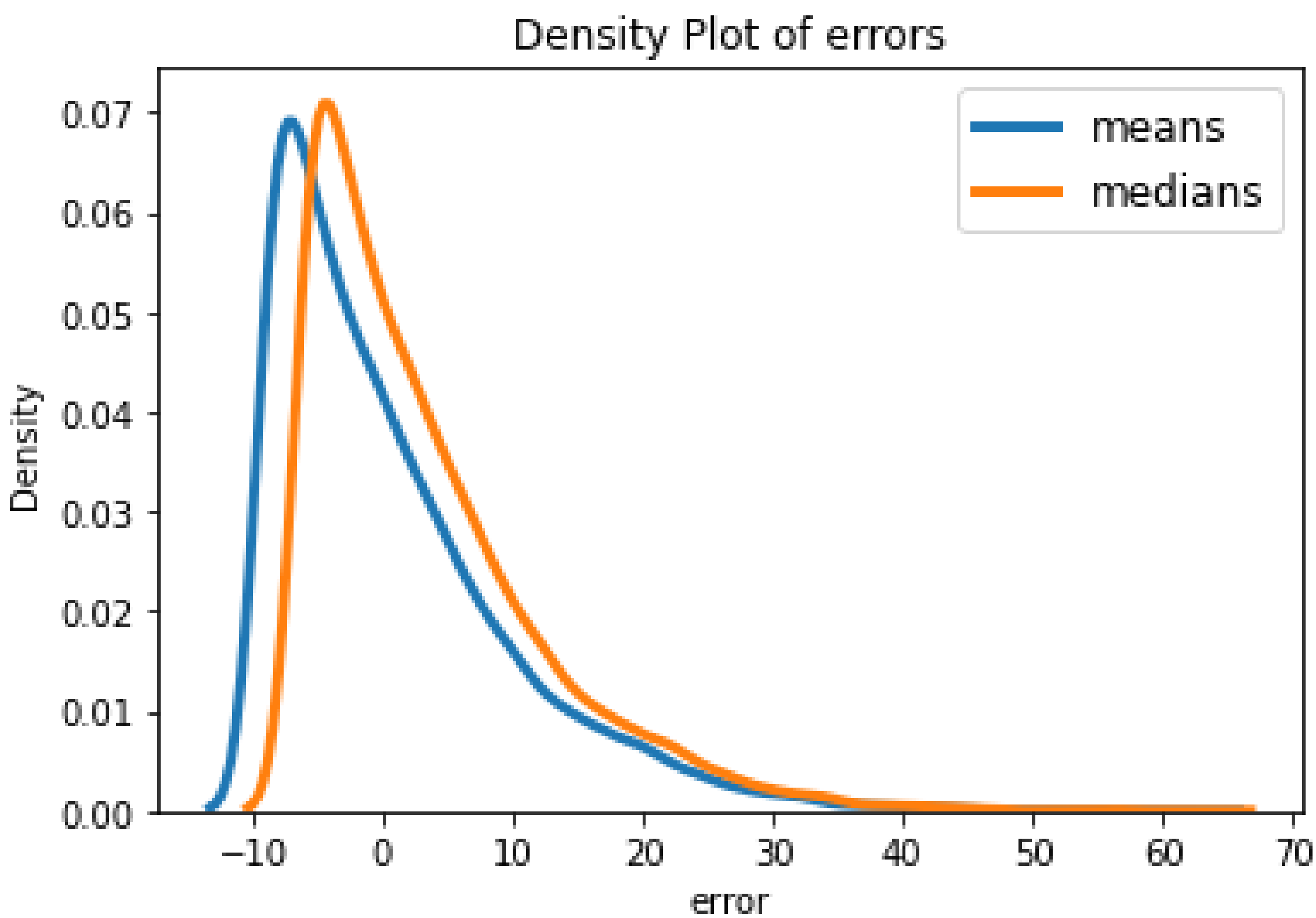
2. Предположим, что вы приезжаете в новый город и видите автобус с номером 100. Требуется с помощью байесовского подхода оценить общее количество автобусных маршрутов в городе. Каким априорным распределением стоит воспользоваться (обоснуйте выбор его параметров)? Какая из статистик апостериорного распределения будет наиболее адекватной (обоснуйте свой выбор)? Как изменятся оценки на количество автобусных маршрутов при последующем наблюдении автобусов с номерами 50 и 150? *Подсказка: воспользоваться результатами предыдущей задачи. При этом обдумать как применить непрерывное распределение к дискретным автобусам.*

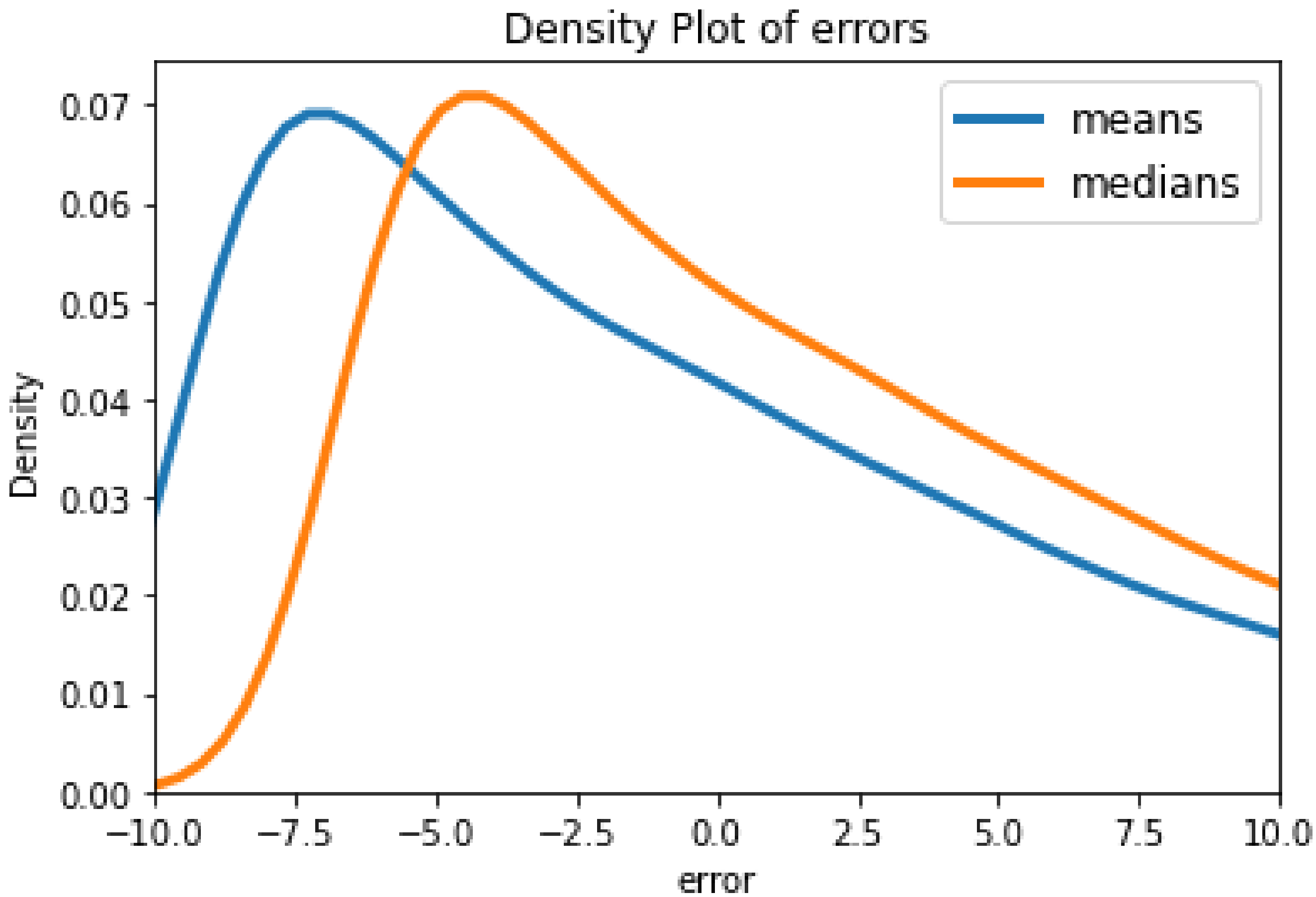
   Опыт городского жителя подсказывает мне, что обычно автобусы приезжают равновероятно, из чего я делаю вывод, что номер автобуса – величина из равномерного распределения. Будем считать, что номер автобуса – это округленное до целого случайное число из равномерного распепределения.

   Тогда в качестве априорного удобно взять распределение Парето. Мы знаем, что параметры распределения, с приходом нового значения x изменяться следующим образом: $\hat{\alpha} = 1 + \alpha; \hat{\beta} = \max(\beta, x)$. При таком обновлении логично выбрать параметры $\alpha = 1$ – так как мы видели один автобус и $\beta = 100$ – так как это номер автобуса, который мы пронаблюдали.

   Нетрудно показать, что для $x > 1$: $2^{\frac{1}{x}} < \left(1 + \frac{1}{x}\right)$ (функции монотонны и пересекаются в единице). Значит оценка матожиданием всегда больше мединаны, которая в свою очередь всегда больше моды. Я был уверен, что стоит быть пессимстом и считать, что человек – существо невезучее, а значит надо выбирать матожидание как самую верхнюю из доступных оценок. Но потом провел эксперименты. Представим, что мы попали в город, где 100 автобусов, увидев всего 10 мы оценили наш параметр матожиданием и медианой. Записали значения. А потом все повторилось 10000 раз. Мы достали нашу тетрадку с записями и посмотрели на распределение ошибки (то есть 100 минус наша оценка).

   Ядерные оценки полученных распределений представлены на рисунках ниже.





И на взгляд, и статистически видно, что оценивая медианой ответы получаются более близкие правильному результату. Оданако если наложить дополнительные ограничения, например бОльший штраф за положительную ошибку, то оценка средним станет предпочтительнее.

Как и говорилось выше, оценка матожиданием всегда будет больше оценки мединой и модой. Когда мы пронаблюдаем автобус 150 все оценки вырастут, когда пронаблюдаем 50 уменьшаться только оценки медианой и модой (так как увеличится количество автобусов, которые мы видели, но не изменится максимальный номер автобуса).

3. Записать распределение Парето с плотностью $\text{Pareto}(x|a,b) = \frac{ba^b}{x^{b+1}}[x \geq a]$ при фиксированном a в форме экспоненциального класса распределений. Найти $\mathbb{E}\log x$ путём дифференцирования нормировочной константы.

Запишем определение экспоненциального класса: $p(x|\alpha) = \frac{f(x)}{g(\alpha)}\exp(\alpha u(x))$
Чтобы "забороть"экспоненту, нам придется брать логарифм чего-то, что зависит от x и входит в искомую плотность месте с α. Возьмем $u(x) = -\ln x$, тогда $\exp -\alpha \ln x = \frac{1}{x^\alpha}$
Положив $f(x) = \dfrac{[x \geq \beta]}{x}$ и $g(\alpha) = \dfrac{1}{\alpha\beta^\alpha}$, получим что в экспоненциальной форме искомое распределение (при фиксированном β) выглядит как $p(x|\alpha,\beta) = \dfrac{\alpha\beta^\alpha}{x}\exp(-\alpha\ln x)[x \geq \beta]$
$\mathbb{E}\ln x = -\dfrac{\partial g(\alpha)}{\partial\alpha} = \dfrac{1}{\alpha^2\beta^{2\alpha}}\left(\beta^\alpha + \alpha\beta^\alpha\ln\beta\right) = \dfrac{1+\alpha\ln\beta}{\alpha^2\beta}$