...

Цель этой заметки — показать связь мультиномильной модели с пуассоновским потоком, а заодно доказать формулу для вероятностей :)

Равномерное, экспоненциальное и распределение Гумбеля

Величина L распределена равномерно на отрезке $[0;1]$. Посмотрим на цепочку

$$L \xrightarrow{h} M \xrightarrow{h} R, \quad h(x) = -\ln x,$$

где каждая следующая величина получается как минус логарифм предыдущей, $M = -\ln L$, $R = -\ln M$. В обратную сторону, $M = \exp(-R)$, $L = \exp(-M)$,

$$L \xleftarrow{g} M \xleftarrow{g} R, \quad g(x) = \exp(-x).$$

Найдём, как меняется функция плотности через дифференциальную форму. Стартуем с плотности $f_L(\ell) = 1$ или с формы $f_L(\ell)d\ell = 1 \cdot d\ell$. И просто подставляем обратное преобразование $\ell = \exp(-m)$ в $f_L(\ell)d\ell$,

$$f_L(\ell)d\ell = 1 \cdot d\ell = 1 \cdot d(\exp(-m)) = -\exp(-m)dm.$$

Отсюда, $f_M(m) = \exp(-m)$, случайная величина M имеет *экспоненциальное распределение* с единичной интенсивностью. Знак минус вовсе не говорит, что плотность отрицательная :) Дело лишь в том, что положительное $d\ell$ соответствует отрицательному dm.

Теперь подставляем обратное преобразование $m = \exp(-r)$ в $f_M(m)dm$,

$$f_M(m)dm = \exp(-m) \cdot dm = \exp(-\exp(-r)) \cdot d(\exp(-r)) = -\exp(-r - \exp(-r))dr.$$

Замечаем, что функция плотности величины R равна $f_R(r) = \exp(-r - \exp(-r))$. Величина R имеет распределение Гумбеля.

Снимем маски с полученной цепочки :)

$$\mathrm{Unif}[0;1] \xrightarrow{h} \mathrm{Expo}(\lambda = 1) \xrightarrow{h} \mathrm{Gumbel}, \quad h(x) = -\ln x.$$

Мультиномиальная модель на трёх языках

Изложим классическую мультиномиальную модель на примере трёх альтернатив на языках трёх распределений.

С помощью распределения Гумбеля:

$$\begin{cases} u_{ij} \sim \mathrm{Gumbel} \\ y_{ia}^* = x_i^T \beta_a + u_{ia}, \text{ где } \beta_a = 0 \\ y_{ib}^* = x_i^T \beta_b + u_{ib} \\ y_{ic}^* = x_i^T \beta_c + u_{ic} \\ y_i = \begin{cases} a, \text{ если } y_{ia}^* = \max\{y_{ia}^*, y_{ib}^*, y_{ic}^*\} \\ b, \text{ если } y_{ib}^* = \max\{y_{ia}^*, y_{ib}^*, y_{ic}^*\} \\ c, \text{ если } y_{ic}^* = \max\{y_{ia}^*, y_{ib}^*, y_{ic}^*\} \end{cases} \end{cases}$$

С помощью экспоненциального распределения, $v_{ij} = \exp(-u_{ij})$, $T_{ij} = \exp(-y^*_{ij})$.

$$\begin{cases} v_{ij} \sim \text{Expo}(\lambda = 1) \\ T_{ia} = v_{ia} / \exp(x_i^T \beta_a), \text{ где } \beta_a = 0 \\ T_{ib} = v_{ib} / \exp(x_i^T \beta_b) \\ T_{ic} = v_{ic} / \exp(x_i^T \beta_c) \\ y_i = \begin{cases} a, \text{ если } T_{ia} = \min\{T_{ia}, T_{ib}, T_{ic}\} \\ b, \text{ если } T_{ib} = \min\{T_{ia}, T_{ib}, T_{ic}\} \\ c, \text{ если } T_{ic} = \min\{T_{ia}, T_{ib}, T_{ic}\} \end{cases} \end{cases}$$

Замечаем, что в экспоненциальном изложении мультиномиальной логит-модели $T_{ij} \sim \text{Expo}(\lambda = \exp(x_i^T \beta_j))$.

С помощью равномерного распределения, $w_{ij} = \exp(-v_{ij})$, $S_{ij} = \exp(-T_{ij})$.

$$\begin{cases} w_{ij} \sim \text{Unif}[0, 1] \\ S_{ia} = w_{ia}^{1/\exp(x_i^T \beta_a)}, \text{ где } \beta_a = 0 \\ S_{ib} = w_{ib}^{1/\exp(x_i^T \beta_b)} \\ S_{ic} = w_{ic}^{1/\exp(x_i^T \beta_c)} \\ y_i = \begin{cases} a, \text{ если } S_{ia} = \max\{S_{ia}, S_{ib}, S_{ic}\} \\ b, \text{ если } S_{ib} = \max\{S_{ia}, S_{ib}, S_{ic}\} \\ c, \text{ если } S_{ic} = \max\{S_{ia}, S_{ib}, S_{ic}\} \end{cases} \end{cases}$$

Бинарная логит-модель с помощью Гумбеля и логистического

Бинарную логит модель обычно рассказывают в терминах логистического распределения:

$$\begin{cases} u_i \sim \text{Logistic} \\ y_i^* = x_i^T \beta + u_i \\ y_i = \begin{cases} 0, \text{ если } y_i^* < 0 \\ 1, \text{ если } y_i^* \geq 0 \end{cases} \end{cases}$$

Можно изложить её и в терминах Гумбеля

$$\begin{cases} u_{ij} \sim \text{Gumbel} \\ y_{ia}^* = x_i^T \beta_a + u_{ia}, \text{ где } \beta_a = 0 \\ y_{ib}^* = x_i^T \beta_b + u_{ib} \\ y_i = \begin{cases} a, \text{ если } y_{ia}^* = \max\{y_{ia}^*, y_{ib}^*\} \\ b, \text{ если } y_{ib}^* = \max\{y_{ia}^*, y_{ib}^*\} \end{cases} \end{cases}$$

Замечаем, что изложение в терминах Гумбеля переводится в классическое с логистическим распределением связкой

$$y_i^* = y_{ib}^* - y_{ia}^* = x_i^T \beta_b + u_{ib} - u_{ia}.$$

В узких кругах широко известно, что разница двух независимых распределений Гумбеля $u_{ib} - u_{ia}$ имеет логистическое распределения. Тут пруф :)

Первое событие в конкурирующих потоках

Представим себе дюжину $3 + 4 + 5 = 12$ независимых пуассоновских потоках единичной интенсивности. На первые три потока поставим букву A, на следующие четыре потока — букву B, на следующие пять потоков — букву C.

Зададимся вопросом, на какую букву придётся самое первое происшествие в этих 12 потоках? В силу симметрии, первое происшествие ложится равновероятно на каждый исходный из дюжины потоков. Поэтому ответ имеет простой вид

$$\mathbb{P}(A) = \frac{3}{3+4+5}, \quad \mathbb{P}(B) = \frac{4}{3+4+5}, \quad \mathbb{P}(C) = \frac{5}{3+4+5}.$$

Теперь обратим внимание, что A-потоки можно сложить и получить один поток с интенсивностью три происшествия в час, B-потоки можно сложить в один поток с интенсивностью четыре, и C-потоки можно сложить в один поток с интенсивностью пять. Свойство суммирования для пуассоновского потока можно доказать, например, сложив независимые пуассоновские случайные величины. Также можно построить доказательство, опираясь на аксиомы пуассоновского потока и математическое ожидание числа событий за единицу времени.

И мы видим следующий факт:

Если k независимых пуассоновских потоков имеют интенсивности λ_1, λ_2, ..., λ_k, а T_j — время первого происшествия в j-м потоке, то

$$\mathbb{P}(T_j = \min\{T_1, T_2, \ldots, T_k\}) = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_k}.$$

Другими словами, вероятность того, что на j-й поток придётся первое происшествие пропорциональна его интенсивности.

Мультиномиальная логит-модель

Рассмотрим некоторого индивида и три альтернативы, A, B и C. Представим себе мысленный эксперимент. Индивид изготавливает три кастрюли в которые капли дождя попадают с интенсивностями $\lambda_A = \exp(V_A)$, $\lambda_B = \exp(V_B)$, $\lambda_C = \exp(V_C)$. Можно считать, что λ_j — это площадь поверхности кастрюли, а индивид выберет ту альтернативу, куда попадёт первая капля дождя.

Поток дождя, попадающий в каждую из кастрюль — это пуассоновский поток :)

Вероятности выбора альтернатив равны

$$\mathbb{P}(A) = \frac{\lambda_A}{\lambda_A + \lambda_B + \lambda_C}, \ldots$$

Время попадания первой капли в j-ю кастрюлю T_j имеет экспоненциальное распределение с параметром λ_j. Отсюда, $\lambda_j T_j \sim \text{Expo}(1)$ и $u_j = -\ln(\lambda_j T_j) \sim \text{Gumbel}$,

$$u_j = -V_j - \ln T_j \sim \text{Gumbel}$$

Назовём величину $y_j^* = -\ln T_j$ полезностью от альтернативы j,

$$y_j^* = -\ln T_j = V_j + u_j$$

Индивид выбирает ту альтернативу, у которой полезность y_j^* больше! По-доказанному,

$$\begin{aligned}\mathbb{P}(y_j^* = \max\{y_1^*, \ldots, y_k^*\}) = \mathbb{P}(T_j = \min\{T_1, T_2, \ldots, T_k\}) = \\ = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_k} = \frac{\exp(V_j)}{\exp(V_1) + \exp(V_2) + \cdots + \exp(V_k)}\end{aligned}$$

Классическое доказательство

Для полноты картины приведём классическое доказательство формулы для вероятностей в мультиномиальной логит-модели напрямую из распределения Гумбеля. ...

1. Источники мудрости