Математические основы искусственного интеллекта
Системы случайных величин

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

Ноябрь 2021

Мотивировка и план лекции

Ранее мы изучили распределение случайной величины, функцию распределения.

В реальных задачах как правило приходится изучать не одну случайную величину, а несколько. Это может быть семемйство случайных величин или их последовательность.

Изучая совместное поведение нескольких случайных величин, мы сталкиваемся с принципиально новыми явлениями и сложностями.

1. Функция совместного распределения
2. Независимость случайных величин

Системы случайных величин

Пусть случайные величины $\xi_1, \ldots, \xi_n$ заданы на одном вероятностном пространстве $\langle \Omega, \mathcal{F}, \mathrm{P} \rangle$.

Определение

Функция $F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \mathrm{P}(\xi_1 < x_1, \ldots, \xi_n < x_n)$ называется функцией распределения вектора $(\xi_1, \ldots, \xi_n)$ или функцией совместного распределения случайных величин $\xi_1, \ldots, \xi_n$.

Распределение двумерной дискретной случайной величины

Распределением двумерной дискретной случайной величины (ξ, η) называется перечень всех возможных пар (x_i, y_j) и соответствующих вероятностей $p_{ij} = P(\xi = x_i,\, \eta = y_j)$.

$\xi \mid \eta$	y_1	y_2	$\dots$	y_m
x_1	p_{11}	p_{12}	$\dots$	p_{1m}
x_2	p_{21}	p_{22}	$\dots$	p_{2m}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_n	p_{n1}	p_{n2}	$\dots$	p_{nm}

Зная распределение двумерной случайной величины, можно найти распределение ее составляющих.

$$P(\xi = x_1) = p_{11} + p_{12} + \ldots + p_{1m}.$$

Распределение двумерной дискретной случайной величины

Распределением двумерной дискретной случайной величины (ξ, η) называется перечень всех возможных пар (x_i, y_j) и соответствующих вероятностей $p_{ij} = P(\xi = x_i,\ \eta = y_j)$.

$\xi \mid \eta$	y_1	y_2	y_3
x_1	0.10	0.30	0.20
x_2	0.06	0.18	0.16

Зная распределение двумерной случайной величины (ξ, η), найти распределение ее составляющих.

η	y_1	y_2	y_3
P	0.16	0.48	0.36

Пример совместного дискретного распределения

Миша и Катя самозанятые и ведут совместный бюджет, каждый работает независимо от другого. Заработок Миши — случайная величина распределенная по закону

Заработок Миши	0	100	500
P	0.1	0.7	0.2

Заработок Кати — тоже случайная величина распределенная по закону

Заработок Кати	0	200	400
P	0.3	0.5	0.2

Составьте закон совместного распределения.

Совместный	0	100	200	300	400	500	700	900
P	0.03	0.21	0.05	0.35	0.02	0.20	0.10	0.04

Пример совместного дискретного распределения

Решение, представленное на предыдущем слайде, неверное!

Заработок Миши	0	100	500
P	0.1	0.7	0.2

Заработок Кати	0	200	400
P	0.3	0.5	0.2

Закон совместного распределения.

Миша | Катя	0	200	400
0	0.03	0.05	0.02
100	0.21	0.35	0.14
500	0.06	0.10	0.04

Функция распределения дискретной двумерной случайной величины

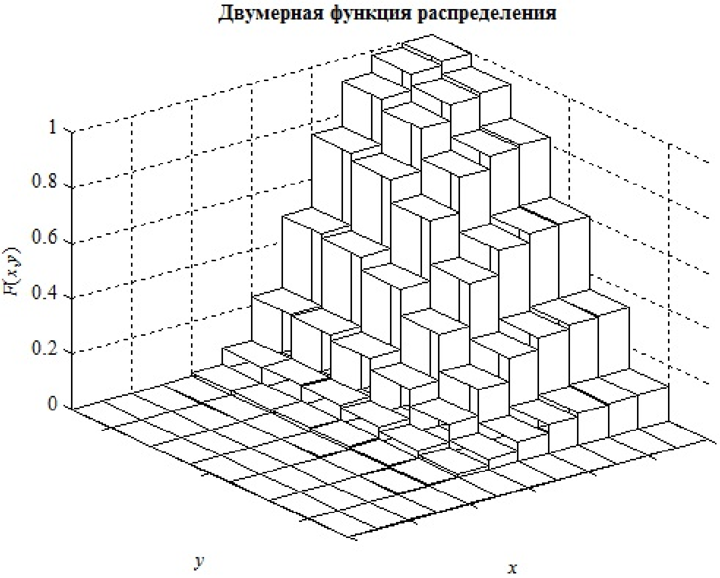
Закон совместного распределения: $P(\xi_1 = x_1, \xi_2 = x_2)$.

Миша | Катя	0	200	400
0	0.03	0.05	0.02
100	0.21	0.35	0.14
500	0.06	0.10	0.04

Функциея распределения: $F_{\xi_1,\xi_2}(x_1,\ x_2) = P(\xi_1 < x_1, \xi_2 < x_2)$

Миша | Катя	$x_2 \leq 0$	$0 < x_2 \leq 200$	$200 < x_2 \leq 400$	$400 < x_2$
$x_1 \leq 0$	0	0	0	0
$0 < x_1 \leq 100$	0	0.03	0.08	0.10
$100 < x_1 \leq 500$	0	0.24	0.64	0.80
$500 < x_1$	0	0.30	0.80	1

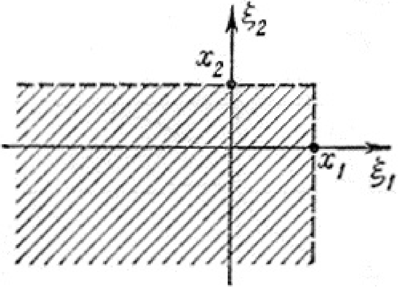
График функция распределения



Двумерная функция распределения

Определение

Функция $F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \mathrm{P}(\xi_1 < x_1, \ldots, \xi_n < x_n)$ называется функцией распределения вектора $(\xi_1, \ldots, \xi_n)$ или функцией совместного распределения случайных величин $\xi_1, \ldots, \xi_n$.



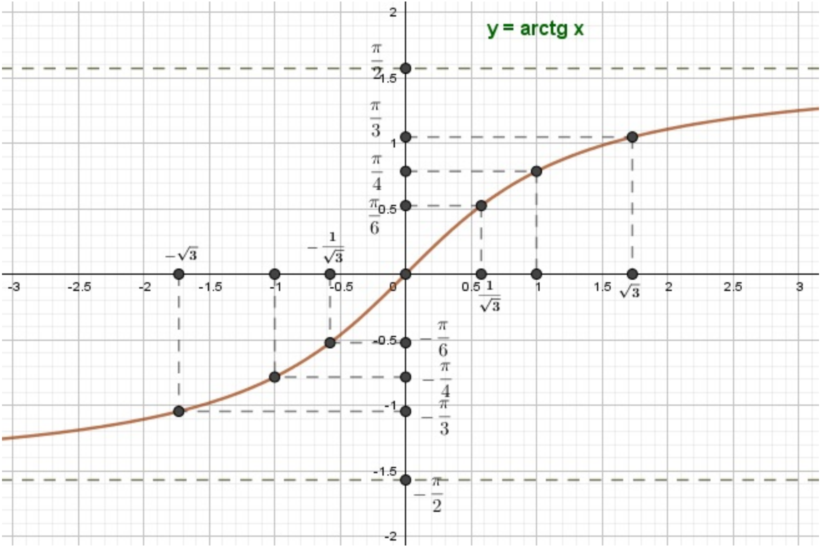
Функция распределения непрерывной двумерной случайной величины

Двумерная случайная величина задана функцией распределения

$$F_{\xi_1,\,\xi_2}(x_1,\; x_2) = \left(\frac{1}{\pi}\arctan x + \frac{1}{2}\right) \times \frac{e^{x_2}}{1+e^{x_2}}$$

Найти вероятность того, что в результате испытания составляющая ξ_1 примет значение меньше 1, а составляющая ξ_2 примет значение меньше 0.

График арктангенса



Зачем нам нужен арктангенс?

Может показаться, что рассмотренный пример с арктангенсом и экспонентами оторван от жизни и к реальным задачам не может иметь отношения... Однако, это совершенно не так!

Во многих задачах возникает необходимоть по наблюдаемым данным предсказывать бинарный отклик: отсдаст заемщик кредит или нет, продолжит подписку клиент или откажется, умрет пациент или останется жив — вот неполный список подобных задач. Наблюдаемыми данными в случае кредитного соринга могут быть возраст заемщика, его месячный доход и величины долгов по кредитным картам.

Зачем нам нужен арктангенс?

В таких задачах строят модель для предсказания вероятности, а значит, использовать модель линейной регрессии здесь нельзя, иначе при определенных значениях предикторов предсказываемая вероятность окажется больше единицы. Необходимо использовать такое нелинейное преобразование, что вычисляемые значения будут принадлежать промежутку $[0, 1]$.

Как раз функции

$$\left(\frac{1}{\pi}\arctan x + \frac{1}{2}\right)$$

$$\frac{e^{x_2}}{1 + e^{x_2}}$$

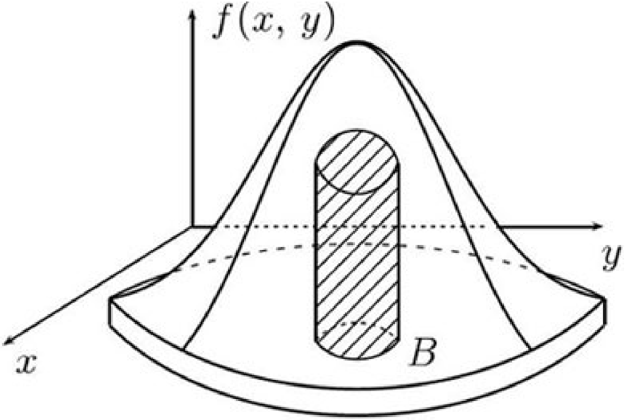
удовлетворяют нужным свойствам и могут использоваться в задачах кредитного скоринга и других, где нужно моделировать бинарный отклик.

Свойства функции совместного распределения

Для простоты обозначений ограничимся вектором (ξ_1, ξ_2) из двух величин.
(F0) Для любых x_1, x_2 верно неравенство: $0 \le F_{\xi_1,\xi_2}(x_1, x_2) \le 1$.
(F1) $F_{\xi_1,\xi_2}(x_1, x_2)$ не убывает по каждой координате вектора (x_1, x_2).
(F2) Для любого $i = 1, 2$ существует $\lim\limits_{x_i \to -\infty} F_{\xi_1,\xi_2}(x_1, x_2) = 0$.
Существует двойной предел $\lim\limits_{x_1 \to +\infty} \lim\limits_{x_2 \to +\infty} F_{\xi_1,\xi_2}(x_1, x_2) = 1$.
(F3) Функция $F_{\xi_1,\xi_2}(x_1, x_2)$ по каждой координате вектора (x_1, x_2) непрерывна слева.
(F4) Чтобы по функции совместного распределения восстановить функции распределения ξ_1 и ξ_2 в отдельности, следует устремить мешающую переменную к $+\infty$:

$$\lim_{x_1 \to +\infty} F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_2}(x_2), \qquad \lim_{x_2 \to +\infty} F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_1}(x_1).$$

Плотность распределения



Плотность распределения

Опрпеделение

Случайные величины ξ_1, ξ_2 имеют абсолютно непрерывное совместное распределение, если существует неотрицательная функция $f_{\xi_1,\,\xi_2}(x,\, y)$ такая, что для любого множества $B \in \mathfrak{B}(\mathbb{R}^2)$ имеет место равенство

$$\mathrm{P}((\xi_1,\, \xi_2) \in B) = \int\!\!\int\limits_B f_{\xi_1,\,\xi_2}(x,\, y)\, dx\, dy.$$

Плотность распределения

Если случайные величины ξ_1, ξ_2 имеют абсолютно непрерывное совместное распределение, то для любых x_1, x_2 имеет место равенство

$$F_{\xi_1,\,\xi_2}(x_1,\,x_2) = \mathrm{P}(\xi_1 < x_1,\,\xi_2 < x_2) = \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{x_2} f_{\xi_1,\,\xi_2}(x,\,y)\,dy \right) dx.$$

Если совместное распределение абсолютно непрерывно, то по функции совместного распределения его плотность находится как смешанная частная производная:

$$f_{\xi_1,\,\xi_2}(x,\,y) = \frac{\partial^2}{\partial x \partial y} F_{\xi_1,\,\xi_2}(x,\,y).$$

Задачи формальные

1. Случайная величина ξ задана плотностью распределения

$$f_\xi(x) = \begin{cases} 0, & \text{если } x < 0, \\ a \sin x, & \text{если } 0 \le x \le \pi, \\ 0, & \text{если } \pi < x. \end{cases} \tag{1}$$

   Найти параметр a, функцию распределения, вероятность попадания в интервал $[0, \pi/4]$, математическое ожидание.

2. Найти вероятность попадания случайной точки (ξ, η) в прямоугольник $[\pi/6,\ \pi/2] \times [\pi/4,\ \pi/3]$, если известна функция распределения

$$F_{\xi,\eta}(x, y) = \sin x \sin y, \quad (x, y) \in [0,\ \pi/2] \times [0,\ \pi/2].$$

   В условии задачи функция $F_{\xi,\eta}(x, y)$ задана не на всем множестве. А как надо формально продлить ее на $\mathbb{R}^2$?
   Найти плотность распределения $f_{\xi,\eta}(x, y)$.

Задачи формальные

3. Плотность распределения двумерной случайной величины задана следующим образом

$$f_{\xi,\eta}(x,y) = \begin{cases} C\cos x\cos y, & \text{если } (x,y) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \\ 0, & \text{иначе }. \end{cases} \tag{2}$$

Найти напамерт C, вероятность попадания в прямоугольник $\left[0, \frac{\pi}{6}\right] \times \left[0, \frac{\pi}{3}\right]$, функцию распределения $F_{\xi,\eta}(x,y)$. Найти плотности распределений $f_{\xi}(x)$ и $f_{\eta}(y)$.