Теория вероятностей. Лекция девятая
Условное математическое ожидание

Дмитрий Валерьевич Хлопин
glukanat@mail.ru

Институт математики и механики им. Н.Н.Красовского

30.10.2018

Что дальше?

- распределение случайных величин
- медиана и математическое ожидание
- независимость случайных величин
- производящие функции как матожидание
- дисперсия, ковариация и корреляция
- векторные случайные величины, их совместные распределения, маргинальные распределения
- энтропия и условная энтропия
- условное математическое ожидание

Что сделано:
мы научились находить линейную (как и любую наперед заданную двухпараметрическую) зависимость значения одной случайной величины с помощью значений другой случайной величины. Для этого требуется рассчитать их дисперсии и ковариации, для чего достаточно знать матожидания.
С совместным распределением можно больше...

Напомним: случайная величина, ее распределение и ее энтропия

В вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ дискретная случайная величина $\xi : \Omega \to \mathbb{R}$ задает распределение

x_1	x_2	x_3	x_4	$\dots$
p_1	p_2	p_2	p_4	$\dots$

Здесь $p_i \overset{\triangle}{=} \mathbb{P}(\xi = x_i) = \mathbb{P}\{\omega \in \Omega \,|\, \xi(\omega) = x_i\}$.

Число $H(\xi) \overset{\triangle}{=} -\sum_i p_i \log p_i$ называют энтропией этого распределения (этой случайной величины).

Напомним: совместные распределения

В вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ векторная дискретная случайная величина $\vec{\xi} = (Y, X) : \Omega \to \mathbb{R}^2$ задает совместное распределение

	x_1	x_2	x_3	$\ldots$
y_1	$p_{11} \triangleq \mathbb{P}(\vec{\xi} = (y_1, x_1))$	$p_{12} \triangleq \mathbb{P}(\vec{\xi} = (y_1, x_2))$	$p_{13} \triangleq \mathbb{P}(\vec{\xi} = (y_1, x_3))$	$\ldots$
y_2	$p_{21} \triangleq \mathbb{P}(\vec{\xi} = (y_2, x_1))$	$p_{22} \triangleq \mathbb{P}(\vec{\xi} = (y_2, x_2))$	$p_{23} \triangleq \mathbb{P}(\vec{\xi} = (y_2, x_3))$	$\ldots$
y_3	$p_{31} \triangleq \mathbb{P}(\vec{\xi} = (y_3, x_1))$	$p_{32} \triangleq \mathbb{P}(\vec{\xi} = (y_3, x_2))$	$p_{33} \triangleq \mathbb{P}(\vec{\xi} = (y_3, x_3))$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

У такой случайной величины и ее совместного распределения энтропия считается по формуле

$$H(\vec{\xi}) \triangleq -\sum_{i,j} p_{ij} \log p_{ij}.$$

Напомним: свойства энтропии

1^0 $H(X) \geq 0$, причем $H(X) = 0$ тогда и только тогда, когда X — вырожденная случайная величина;

2^0 $H(f(X)) \leq H(X)$ для любой функции $f:\mathbb{R}\to\mathbb{R}$,

причем $H(f(X)) = H(X)$ тогда и только тогда, когда f на $\{x_1, x_2, \ldots\}$ — биекция;

3^0 $H(X) \leq \log N$, если X принимает не более чем N значений,

причем $H(X) = \log N$ только при $p_i = 1/N$;

4^0 для $\vec{\xi} = (X, Y)$ выполнено $H(\vec{\xi}) \leq H(X) + H(Y)$, причем $H(\vec{\xi}) = H(X) + H(Y)$ в точности для независимых X, Y.

Напомним: условная энтропия через условное распределение

$$H(X|Y) \triangleq H(\vec{\xi}) - H(Y) = -\sum_j q_j \sum_i \frac{p_{ij}}{q_j} \log \frac{p_{ij}}{q_j}$$
$$= -\sum_j q_j \sum_i \mathbb{P}(X = x_i | Y = y_j) \log \mathbb{P}(X = x_i | Y = y_j),$$

где $\mathbb{P}(X|Y)$ — условное распределение случайной величины X относительно Y:

при	x_1	x_2	x_3	$\ldots$
$y_1:$	$\mathbb{P}(X = x_1 \mid Y = y_1) = \frac{p_{11}}{q_1}$	$\mathbb{P}(X = x_2 \mid Y = y_1) = \frac{p_{12}}{q_1}$	$\mathbb{P}(X = x_3 \mid Y = y_1) = \frac{p_{13}}{q_1}$	$\ldots$
$y_2:$	$\mathbb{P}(X = x_1 \mid Y = y_2) = \frac{p_{21}}{q_2}$	$\mathbb{P}(X = x_2 \mid Y = y_2) = \frac{p_{22}}{q_2}$	$\mathbb{P}(X = x_3 \mid Y = y_2) = \frac{p_{23}}{q_2}$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

Как восстановить X по Y, если условная энтропия $H(X|Y)$ равна нулю...

Подумать: X является функцией от Y, то есть $X = f(Y)$, тогда и только тогда, когда

$$H(X|Y) = -\sum_j q_j \sum_i \mathbb{P}(X = x_i|Y = y_j) \log \mathbb{P}(X = x_i|Y = y_j) = 0.$$

при	x_1	x_2	x_3	$\dots$
$y_1:$	$\mathbb{P}(X=x_1\|Y=y_1)=\frac{p_{11}}{q_1}$	$\mathbb{P}(X=x_2\|Y=y_1)=\frac{p_{12}}{q_1}$	$\mathbb{P}(X=x_3\|Y=y_1)=\frac{p_{13}}{q_1}$	$\dots$
$y_2:$	$\mathbb{P}(X=x_1\|Y=y_2)=\frac{p_{21}}{q_2}$	$\mathbb{P}(X=x_2\|Y=y_2)=\frac{p_{22}}{q_2}$	$\mathbb{P}(X=x_3\|Y=y_2)=\frac{p_{23}}{q_2}$	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

Подумать: вроде бы X является (или не является) функцией от Y вне зависимости от выбора вероятности $\mathbb{P}$...

Измеримость случайных величин: банальности и не только

Любая дискретная случайная величина ξ порождает σ-алгебру $\sigma(\xi)$, для дискретных случайных величин ее можно выразить формулой

$$\sigma(\xi) \triangleq \sigma\Big\{\{\omega|\xi(\omega) = x\}\big|x \in \mathbb{R}\Big\}$$

(это обозначение уже встречалось ранее при определении независимости случайных величин).

Если $\sigma(\xi) \subset \mathcal{H}$ для некоторой σ-алгебры $\mathcal{H}$, то говорят, что случайная величина ξ $\mathcal{H}$-измерима (измерима относительно σ-алгебры $\mathcal{H}$).

Для любой σ-алгебры $\mathcal{H}$ выполнено $\{\Omega, \varnothing\} \subset \mathcal{H} \subset \mathcal{F}$;

$\Rightarrow$ любая дискретная случайная величина $\mathcal{F}$-измерима;

$\Rightarrow$ константы, и только они, $\{\Omega, \varnothing\}$-измеримы.

Если $X = f(Y)$, то X — $\sigma(Y)$-измеримая случайная величина.

Как проверить, является ли X функцией от Y, без применения вероятности...

Теорема. [с-но; от 0,5 до 1 балла] X — $\sigma(Y)$-измеримая случайная величина тогда и только тогда, когда $X = f(Y)$ для некоторой функции $f : \mathbb{R} \to \mathbb{R}$.

при	x_1	x_2	x_3	$\dots$
y_1 :	$\mathbb{P}(X=x_1 \vert Y=y_1) = \frac{p_{11}}{q_1}$	$\mathbb{P}(X=x_2 \vert Y=y_1) = \frac{p_{12}}{q_1}$	$\mathbb{P}(X=x_3 \vert Y=y_1) = \frac{p_{13}}{q_1}$	$\dots$
y_2 :	$\mathbb{P}(X=x_1 \vert Y=y_2) = \frac{p_{21}}{q_2}$	$\mathbb{P}(X=x_2 \vert Y=y_2) = \frac{p_{22}}{q_2}$	$\mathbb{P}(X=x_3 \vert Y=y_2) = \frac{p_{23}}{q_2}$	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

Что делать, если X не является $\sigma(Y)$-измеримой,
а выразить X через Y надо?
Приближать $\sigma(Y)$-измеримой!!

Напомним: условная вероятность относительно σ-алгебры

Пусть имеются полная группа событий $H_1, \dots, H_k, \dots$ $(\Omega = \bigsqcup_{i=1}^{\infty} H_i)$, например заданных правилом $H_i \overset{\triangle}{=} \{\omega \in \Omega \,|\, Y(\omega) = y_i\}$, и порожденная ей σ-алгебра $\mathcal{H} = \sigma(H_1, \dots, H_k, \dots)$. Пусть $q_i \overset{\triangle}{=} \mathbb{P}(H_i) > 0$.

Условной вероятностью $\mathbb{P}(W|\mathcal{H})$ события $W \in \mathcal{F}$ относительно σ-алгебры $\mathcal{H} = \sigma(H_1, \dots, H_k, \dots)$ называют $\mathcal{H}$-измеримую функцию

$$\Omega \ni \omega \mapsto \mathbb{P}(W|\mathcal{H})(\omega) \overset{\triangle}{=} \begin{cases} \mathbb{P}(W|H_1), & \text{если } \omega \in H_1; \\ \dots \\ \mathbb{P}(W|H_k), & \text{если } \omega \in H_k; \\ \dots \end{cases}$$

Условное матожидание относительно σ-алгебры, I

Пусть имеются полная группа событий $H_1,\dots,H_k,\dots$ $(\Omega = \bigsqcup_{i=1}^{\infty} H_i)$, например заданных правилом $H_i \overset{\triangle}{=} \{\omega \in \Omega \,|\, Y(\omega) = y_i\}$, и порожденная ей σ-алгебра $\mathcal{H} = \sigma(H_1,\dots,H_k,\dots)$. Пусть $q_i \overset{\triangle}{=} \mathbb{P}(H_i) > 0$.

Условным математическим ожиданием $\mathbb{E}(\xi|\mathcal{H})$ дискретной случайной величины ξ относительно σ-алгебры $\mathcal{H} = \sigma(H_1,\dots,H_k,\dots)$ называют отображение $\mathbb{E}(\xi|\mathcal{H}) : \Omega \to \mathbb{R}$, действующее по правилу:

$$\Omega \ni \omega \mapsto \mathbb{E}(\xi|\mathcal{H})(\omega) \overset{\triangle}{=} \begin{cases} \sum_{i\in\mathbb{N}} x_i \mathbb{P}(\xi = x_i|H_1), & \text{если } \omega \in H_1; \\ \dots \\ \sum_{i\in\mathbb{N}} x_i \mathbb{P}(\xi = x_i|H_k), & \text{если } \omega \in H_k; \\ \dots \end{cases}$$

будем говорить, что условное матожидание существует, если, конечно, все выражения выше определены и конечны.

Условное матожидание относительно σ-алгебры, II

Важное замечание: как и условная вероятность $\mathbb{P}(\xi|\mathcal{H})$, условное матожидание $\mathbb{E}(\xi|\mathcal{H})$ является не числом, а $\mathcal{H}$-измеримой функцией, отображающей каждый исход ω в некоторое число.

Условным математическим ожиданием $\mathbb{E}(\xi|\mathcal{H})$ дискретной случайной величины ξ относительно σ-алгебры $\mathcal{H} = \sigma(H_1, \ldots, H_k, \ldots)$ $(\Omega = \bigsqcup_{i=1}^{\infty} H_i)$ называют отображение

$$\Omega \ni \omega \mapsto \mathbb{E}(\xi|\mathcal{H})(\omega) \triangleq \begin{cases} \sum_{i\in\mathbb{N}} x_i \mathbb{P}(\xi = x_i|H_1), & \text{если } \omega \in H_1; \\ \ldots \\ \sum_{i\in\mathbb{N}} x_i \mathbb{P}(\xi = x_i|H_k), & \text{если } \omega \in H_k; \\ \ldots \end{cases}$$

Будем говорить, что $\mathbb{E}(\xi|\mathcal{H})$ существует, если, конечно, все выражения выше определены и конечны.

Условное матожидание относительно случайной величины

Условное матожидание ξ относительно случайной величины η зададим формулой $\mathbb{E}(\xi|\eta) \equiv \mathbb{E}(\xi|\sigma(\eta))$. Иногда его вводят как функцию от области значений случайной величины η, т.е. как регрессию.

Пример расчета по совместному распределению регрессии h_0

(Y, X)	x_1	x_2	x_3	$\dots$	$\mathbb{P}(Y = y_j)$	$\mathbb{E}(X\|Y)$
y_1	p_{11}	p_{12}	p_{13}	$\dots$	q_1	$h_0(y_1) \triangleq \frac{1}{q_1}\sum_{i=1}^{\infty} x_i p_{1i}$
y_2	p_{21}	p_{22}	p_{23}	$\dots$	q_2	$h_0(y_2) \triangleq \frac{1}{q_2}\sum_{i=1}^{\infty} x_i p_{2i}$
y_3	p_{31}	p_{32}	p_{33}	$\dots$	q_3	$h_0(y_3) \triangleq \frac{1}{q_3}\sum_{i=1}^{\infty} x_i p_{3i}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

Подумать: и в этом, и в предыдущем определении требовалось, чтобы ξ, η были скалярными величинами (отображались в числа). Можно ли вместо $\mathbb{R}$ взять что-то более общее?

Посчитать условное матожидание: индикаторные функции

Рассмотрим произвольное событие $A \in \mathcal{F}$ и его индикаторную функцию 1_A. Пусть имеется также полная группа событий $H_1, \dots, H_k, \dots$ с порожденной ей σ-алгеброй $\mathcal{H}$.
Теперь,

$$\mathbb{E}(1_A|\mathcal{H})(\omega) = \begin{cases} \mathbb{P}(A|H_1), & \text{если } \omega \in H_1; \\ \dots & \\ \mathbb{P}(A|H_k), & \text{если } \omega \in H_k. \\ \dots & \end{cases} = \mathbb{P}(A|\mathcal{H})(\omega).$$

Подумать: выразите $\mathbb{E}\xi$ через $\mathbb{E}(\xi|1_A)$.
Подумать: проверьте формулу $\mathbb{E}(\xi|A) = \mathbb{E}(\xi 1_A)/\mathbb{P}(A)$.

Свойства условного матожидания, стандартные+: [с-но]

Пусть на этом слайде всё существует априори...

0^0 $\mathbb{E}(1_A|\mathcal{H}) = \mathbb{P}(A|\mathcal{H})$ для всех $A \in \mathcal{F}$.

1^0 $\mathbb{E}(\xi|\mathcal{H}) \geq 0$, если $\xi \geq 0$ для всех $\omega \in \Omega$.

2^0 $\mathbb{E}(\xi_1|\mathcal{H}) \geq \mathbb{E}(\xi_2|\mathcal{H})$, если $\xi_1(\omega) \geq \xi_2(\omega)$ для всех $\omega \in \Omega$.

3^0 $\mathbb{E}(c|\mathcal{H}) = c$ для вырожденной случайной величины $c \in \mathbb{R}$;

более того, $\mathbb{E}(\xi|\mathcal{H}) = \xi$ для $\mathcal{H}$-измеримой ξ,

в частности, всегда $\mathbb{E}(f(\xi)|\xi) = f(\xi)$, $\mathbb{E}(\xi|\xi) = \xi$.

4^0 $\mathbb{E}(c\xi|\mathcal{H}) = c\mathbb{E}(\xi|\mathcal{H})$ для $c \in \mathbb{R}$,

более того, $\mathbb{E}(\eta\xi|\mathcal{H}) = \eta\mathbb{E}(\xi|\mathcal{H})$, если η $\mathcal{H}$-измерима.

5^0 $\mathbb{E}(\xi_1|\mathcal{H}) + \mathbb{E}(\xi_2|\mathcal{H}) = \mathbb{E}(\xi_1 + \xi_2|\mathcal{H})$.

6^0 $\eta_1\mathbb{E}(\xi_1|\mathcal{H}) + \eta_2\mathbb{E}(\xi_2|\mathcal{H}) = \mathbb{E}(\eta_1\xi_1 + \eta_2\xi_2|\mathcal{H})$ для $\mathcal{H}$-измеримых η_1, η_2.

7^0 $\mathbb{E}(\xi\eta|\mathcal{H}) = \mathbb{E}(\xi|\mathcal{H})\mathbb{E}(\eta|\mathcal{H})$, если $\xi 1_H, \eta 1_H$ независимы для всех $H \in \mathcal{H}$.

Свойства условного матожидания, существование: [с-но]

8^0 $\mathbb{E}(\mathbb{E}(\eta|\mathcal{G}))$ существует тогда и только тогда, когда существует $\mathbb{E}\eta$, в частности, из существования $\mathbb{E}|\eta|$ следует существование $\mathbb{E}(\eta|\mathcal{G})$.

9^0 $\mathbb{E}(\xi|\mathcal{H})$ существует, если ξ ограничена или принимает конечное число значений.

10^0 $\mathbb{E}(|\xi\eta|\,|\mathcal{H})$ существует, если $\mathbb{E}(\xi^2|\mathcal{H})$ и $\mathbb{E}(\eta^2|\mathcal{H})$ существуют.

11^0 $\mathbb{E}(|\xi|\,|\mathcal{H}), \mathbb{E}(\xi|\mathcal{H})$ существуют, если $\mathbb{E}(\xi^2|\mathcal{H})$ существует.

12^0 $g(\mathbb{E}(\xi|\mathcal{H})) \leq \mathbb{E}(g(\xi)|\mathcal{H})$ для любой выпуклой (вниз) скалярной функции g; при этом, если существует матожидание справа, то и слева тоже существует.

13^0 $\mathbb{E}(\sum_{i=1}^{\infty} \xi_i|\mathcal{H})$ существует и равно $\sum_{i=1}^{\infty} \mathbb{E}(\xi_i|\mathcal{H})$, если конечен ряд $\sum_{i=1}^{\infty} \mathbb{E}(|\xi_i|\,|\mathcal{H})(\omega)$ для всех $\omega \in \Omega$.

Условное матожидание как основание перпендикуляра

18^0 Для случайной величины ξ $(\mathbb{E}\xi^2 < +\infty)$ и случайной величины η среди всех таких $h : \mathbb{R} \to \mathbb{R}$, что $\mathbb{E}(h^2(\eta)) < +\infty$, невязка $\mathbb{E}(\xi - h(\eta))^2$ достигает минимума при $h_0(\eta) = \mathbb{E}(\xi|\eta)$; функция h_0 при этом является регрессией, а реализующееся минимальное значение — условной дисперсией $\mathbb{D}(\xi|\eta)$.

Замечание. Фактически в пространстве всевозможных интегрируемых с квадратом дискретных случайных величин ищется перпендикуляр из ξ на бесконечномерное линейное подпространство всех η-измеримых величин. Полученный результат и называют условным матожиданием ξ относительно η.

Доказательство оптимальности регрессии

Доказательство. Можно положить $\xi = \sum_{i=1}^{\infty} x_i 1_{K_i}$ и $\eta = \sum_{j=1}^{\infty} y_j 1_{H_j}$ для некоторых наборов x_i, y_j, K_i $\left(\bigsqcup_{i=1}^{\infty} K_i = \Omega\right)$ и H_j $\left(\bigsqcup_{j=1}^{\infty} H_j = \Omega\right)$. Тогда мы перебираем числа $z_j = h(y_j)$, минимизируя

$$\begin{aligned}\mathbb{E}(\xi - h(\eta))^2 &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i - z_j)^2\mathbb{P}(H_j \cap K_i)\\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i - z_j)^2 p_{ji}.\end{aligned}$$

В силу $\frac{\partial}{\partial z_j}\mathbb{E}(\xi - h(\eta))^2|_{z_j=h_0(y_j)} = 0$, получаем $\sum_{i=1}^{\infty}(x_i - h_0(y_j))p_{ji} = 0$, то есть

$$h_0(y_j) = \frac{\sum_{i=1}^{\infty} x_i p_{ij}}{\sum_{i=1}^{\infty} p_{ji}} = \frac{\sum_{i=1}^{\infty} x_i \mathbb{P}(\xi = x_i, \eta = y_j)}{\mathbb{P}(\eta = y_j)} = \sum_{i=1}^{\infty} x_i \mathbb{P}(\xi = x_i | \eta = y_j).$$