Теория вероятностей и математическая статистика

Многомерные случайные величины.

Глеб Карпов

ВШБ Бизнес-информатика

Введение

Возвращение в дискретный мир

- Точно так же, как мы можем наблюдать результат одновременного подбрасывания двух монет, мы можем наблюдать результат более чем одной случайной величины одновременно.
- Для описания этого эксперимента мы используем новое понятие - многомерная случайная величина или случайный вектор:
$$X = (X_1, X_2, \dots, X_n), \text{ или в двумерном случае } (X, Y).$$
- Сравните это с элементарным исходом вида $\omega = (\text{состояние}_1, \ \dots, \ \text{состояние}_n)$, как например для одновременного или последовательного подбрасывания монет
- Теперь уникальным **элементарным исходом** будет набор одновременно наблюдаемых значений случайных величин. Например $(x, y) = (10, 2)$, где X - число на кубике с 20 гранями, Y - число на кубике с 6 гранями, бросили их вместе и наблюдаем вместе как единую целую картину.

Совместная функция вероятности

- Нас интересует взаимное поведение этих случайных величин.
- Связаны ли они? Сопровождается ли некоторое значение одной величины более часто некоторым другим значением второй величины? Можем ли мы извлекать больше информации о происходящих случайных экспериментах, когда мы учитываем взаимодействие случайных величин друг с другом?
- Для описания этого взаимодействия, или взаимного поведения, мы используем **совместную функцию вероятности**, т.е.:

	$Y = 1$	$Y = 2$
$X = 1$	$\frac{1}{8}$	$\frac{1}{4}$
$X = 2$	$\frac{1}{8}$	$\frac{1}{2}$

Совместная функция вероятности

- В общем виде:

	$Y = y_1$	...	$Y = y_m$
$X = x_1$	$P(\{X = x_1, Y = y_1\})$	...	$P(\{X = x_1, Y = y_m\})$
...	...	...	...
$X = x_n$	$P(\{X = x_n, Y = y_1\})$	...	$P(\{X = x_n, Y = y_m\})$

- Мы также можем записать $P(\{X = x_i, Y = y_j\})$ как $P_{X,Y}(x_i, y_j)$

Свойства

- $P_{X,Y}(x_i, y_j) \geq 0$, для всех возможных $(x_i,\, y_j)$
- Поскольку $P(\Omega) = 1$, здесь мы имеем $\Omega = \{(x_1, y_1), (x_1, y_2), \ldots, (x_n, y_m)\}$. Из этого следует:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P(\{X = x_i, Y = y_j\}) = 1.$$

Маргинальная функция вероятности

- От англ. marginal. В более старых переводах - частная функция вероятности, частное распределение
- Предположим, нас интересует получение результата $X = x_i$, независимо от того, чему равно Y. Это означает, что все следующие пары являются допустимыми, и они образуют специальное событие:

$$A = \{\text{мы получаем } X = x_i\} = \{(x_i,\, y_1), (x_i,\, y_2), \dots, (x_i,\, y_m)\}$$

- Если затем мы хотим найти вероятность этого события, то по свойству аддитивности:

$$P(\{X = x_i\}) = P_{X,Y}(x_i, y_1) + P_{X,Y}(x_i, y_2) + \dots + P_{X,Y}(x_i, y_m),$$

  т.е. мы складываем вероятности всех элементарных исходов, формирующих это событие.
- Это означает, что используя совместную функцию вероятности, мы можем восстановить собственную функцию вероятности случайной величины, как бы изолируя её отдельно от случайного вектора.
- Раньше мы называли это просто **функцией вероятности**, но теперь нам нужно уточнять, потому что добавляется много новых видов функций вероятности.

Маргинальная функция вероятности

- В общем виде:

$$P(\{X = x_i\}) = \sum_{j=1}^{m} P(\{X = x_i, Y = y_j\}) = \sum_{j=1}^{m} P_{X,Y}(x_i, y_j)$$

$$P(\{Y = y_j\}) = \sum_{i=1}^{n} P(\{X = x_i, Y = y_j\}) = \sum_{i=1}^{n} P_{X,Y}(x_i, y_j)$$

- Другими словами, маргинальные вероятномти являются суммами по соответствующей строке или столбцу в таблице совместного распределения:

	$Y = y_1$	...	$Y = y_m$
$X = x_1$	$P(\{X = x_1, Y = y_1\})$	...	$P(\{X = x_1, Y = y_m\})$
...	...	...	...
$X = x_n$	$P(\{X = x_n, Y = y_1\})$	...	$P(\{X = x_n, Y = y_m\})$

Условная функция вероятности

- В многомерном мире мы также можем исследовать вопросы следующего вида: $P(X = x_i | Y = y_j)$
- Мы называем это еще одной функцией вероятности - **условной функцией вероятности**
- Напомним: условная вероятность - это своего рода параметризованная функция, где событие-условие служит параметром. Функции $P(X|A)$ и $P(X|B)$, $A \neq B$, являются разными функциями!
- Применяя определение условной вероятности:

$$P(X = x_i | Y = y_j) = \frac{P(\{X = x_i\} \cap \{Y = y_j\})}{P(Y = y_j)} = \frac{P(\{X = x_i, Y = y_j\})}{P(Y = y_j)}$$

$$P(Y = y_j | X = x_i) = \frac{P(\{X = x_i\} \cap \{Y = y_j\})}{P(X = x_i)} = \frac{P(\{X = x_i, Y = y_j\})}{P(X = x_i)}$$

Независимость

- Наконец, есть важный разговор о независимости случайных величин. В некотором смысле это даже более важно, чем независимость отдельных событий.
- Напомним. Независимость пары событий:

$$P(A) = P(A|B), \quad P(B) = P(B|A)$$

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \to P(A \cap B) = P(A)P(B).$$

Независимость

- Рассмотрим каждую возможную пару следующих событий:

- $\{X = x_1\}$
- $\{X = x_2\}$
- ...
- $\{X = x_n\}$

- $\{Y = y_1\}$
- $\{Y = y_2\}$
- ...
- $\{Y = y_m\}$

Если каждая пара оказывается независимыми событиями, то мы говорим, что случайные величины X и Y независимы.

$$\forall x_i,\ \forall y_j:\ P\{X = x_i\}P\{Y = y_j\} = P(\{X = x_i\} \cap \{Y = y_j\}) = P(\{X = x_i, Y = y_j\}).$$

Задачи

Совместная функция вероятности X и Y задана следующим образом:

	$Y = 1$	$Y = 2$
$X = 1$	$\frac{1}{8}$	$\frac{1}{4}$
$X = 2$	$\frac{1}{8}$	$\frac{1}{2}$

1. Найдите маргинальные функции вероятности для каждой случайной величины.
2. Вычислите условную функцию вероятности X при условии $Y = i$, $i = 1, 2$.
3. Независимы ли X и Y?

Решение задачи

1. Маргинальные функции вероятности

Для X:

$$P(X=1)=\sum_{j=1}^{2}P(X=1,Y=j)=\frac{1}{8}+\frac{1}{4}=\frac{3}{8}$$

$$P(X=2)=\sum_{j=1}^{2}P(X=2,Y=j)=\frac{1}{8}+\frac{1}{2}=\frac{5}{8}$$

Для Y:

$$P(Y=1)=\sum_{i=1}^{2}P(X=i,Y=1)=\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$$

$$P(Y=2)=\sum_{i=1}^{2}P(X=i,Y=2)=\frac{1}{4}+\frac{1}{2}=\frac{3}{4}$$

Решение задачи (продолжение)

2. Условные функции вероятности

$P(X|Y=1)$:

$$P(X=1|Y=1) = \frac{P(X=1,Y=1)}{P(Y=1)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}$$

$$P(X=2|Y=1) = \frac{P(X=2,Y=1)}{P(Y=1)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}$$

$P(X|Y=2)$:

$$P(X=1|Y=2) = \frac{P(X=1,Y=2)}{P(Y=2)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

$$P(X=2|Y=2) = \frac{P(X=2,Y=2)}{P(Y=2)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

Решение задачи (продолжение)

3. Проверка независимости

Проверим условие независимости для пары:

$$P(X=1, Y=1) = \frac{1}{8} \quad \text{и} \quad P(X=1)P(Y=1) = \frac{3}{8} \cdot \frac{1}{4} = \frac{3}{32} \neq \frac{1}{8}$$

Вывод: X и Y **не независимы**.

Функции от случайных векторов

- Мы обсуждали идею, что функция от случайной величины сама является случайной величиной, т.е. $W = g(X)$.
- Функция может быть от более чем одной переменной, и теперь мы можем получить $W = f(X, Y)$.
- Как и в случае с одной переменной, мы можем захотеть предсказать $E[W] = E[f(X, Y)]$.
- Напомним:

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

$$E[g(X)] = \sum_{i=1}^{n} g(x_i) P(X = x_i)$$

- В новом случае:

$$E[g(X, Y)] = \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j) P(X = x_i, Y = y_j)$$

Пример

Совместная функция вероятности X и Y задана следующим образом:

	$Y = 1$	$Y = 2$
$X = 1$	$\frac{1}{8}$	$\frac{1}{4}$
$X = 2$	$\frac{1}{8}$	$\frac{1}{2}$

Рассмотрим функции $W = XY$, $G = X - Y$. Найдите их функции вероятности и математические ожидания.

Пример

Найти математическое ожидание $E[XY]$

Вариант 1: Через распределение новой случайной величины

Сначала найдем распределение $W = XY$:

- $W = 1$: когда $(X = 1, Y = 1)$, $P(W = 1) = P(X = 1, Y = 1) = \frac{1}{8}$
- $W = 2$: когда $(X = 1, Y = 2)$ или $(X = 2, Y = 1)$, $P(W = 2) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$
- $W = 4$: когда $(X = 2, Y = 2)$, $P(W = 4) = P(X = 2, Y = 2) = \frac{1}{2}$
- Теперь вычислим математическое ожидание:

$$E[XY] = E[W] = 1 \cdot \frac{1}{8} + 2 \cdot \frac{3}{8} + 4 \cdot \frac{1}{2} = \frac{1}{8} + \frac{6}{8} + \frac{4}{2} = \frac{23}{8}$$

Пример

Вариант 2: Используя формулу непосредственного подсчета

$$E[XY] = \sum_{i=1}^{2}\sum_{j=1}^{2} x_i y_j P(X = x_i, Y = y_j)$$

$$\begin{aligned} E[XY] &= 1 \cdot 1 \cdot \frac{1}{8} + 1 \cdot 2 \cdot \frac{1}{4} + 2 \cdot 1 \cdot \frac{1}{8} + 2 \cdot 2 \cdot \frac{1}{2} \\ &= \frac{1}{8} + \frac{2}{4} + \frac{2}{8} + \frac{4}{2} \\ &= \frac{23}{8} \end{aligned}$$

Ответ: $E[XY] = \frac{23}{8} = 2.875$

Пример

Найти математическое ожидание $E[X - Y]$

Вариант 1: Через распределение новой случайной величины

Сначала найдем распределение $G = X - Y$:

- $G = -1$: когда $(X = 1, Y = 2)$, $P(G = -1) = P(X = 1, Y = 2) = \frac{1}{4}$
- $G = 0$: когда $(X = 1, Y = 1)$ и $(X = 2, Y = 2)$, $P(G = 0) = P\left(\{(X = 1, Y = 1), (X = 2, Y = 2)\}\right) = \frac{1}{8} + \frac{1}{2} = \frac{5}{8}$
- $G = 1$: когда $(X = 2, Y = 1)$, $P(G = 1) = P(X = 2, Y = 1) = \frac{1}{8}$
- Теперь вычислим математическое ожидание:

$$E[X - Y] = E[G] = 0 \cdot \frac{5}{8} + (-1) \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} = -\frac{1}{8}$$

Пример

Вариант 2: Используя формулу непосредственного подсчета

$$E[X-Y] = \sum_{i=1}^{2}\sum_{j=1}^{2}(x_i - y_j)P(X = x_i, Y = y_j)$$

$$\begin{aligned} E[X-Y] &= (1-1)\cdot\frac{1}{8} + (1-2)\cdot\frac{1}{4} + (2-1)\cdot\frac{1}{8} + (2-2)\cdot\frac{1}{2} \\ &= 0\cdot\frac{1}{8} + (-1)\cdot\frac{1}{4} + 1\cdot\frac{1}{8} + 0\cdot\frac{1}{2} \\ &= 0 - \frac{1}{4} + \frac{1}{8} + 0 \\ &= -\frac{1}{8} \end{aligned}$$

Ответ: $E[X-Y] = -\frac{1}{8} = -0.125$

Линейная комбинация случайных величин

Математическое ожидание комбинации

- Часто важно анализировать поведение суммы двух или более случайных величин
- В основном мы хотим понять, как ведет себя переменная $T = aX \pm bY$, каковы ее характеристики.
- Линейное свойство математического ожидания: $E[aX \pm bY] = aE[X] \pm bE[Y]$.

$$E[T] = \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i \pm by_j) P(X = x_i, Y = y_j) =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} ax_i P(X = x_i, Y = y_j) \pm \sum_{i=1}^{n} \sum_{j=1}^{m} by_j P(X = x_i, Y = y_j) =$$

$$a \sum_{i=1}^{n} x_i \underbrace{\sum_{j=1}^{m} P(X = x_i, Y = y_j)}_{\text{marginal for } X} \pm b \sum_{j=1}^{m} y_j \underbrace{\sum_{i=1}^{n} P(X = x_i, Y = y_j)}_{\text{marginal for } Y} =$$

$$a \sum_{i=1}^{n} x_i P(X = x_i) \pm b \sum_{j=1}^{m} y_j P(Y = y_j) = aE[X] \pm bE[Y]$$

Линейная комбинация случайных величин

Но дисперсия - это уже совсем другая история...

- Нас интересует $Var[T] = Var[aX \pm bY]$.
- Давайте применим сокращенную формулу для дисперсии:

$$\begin{aligned} Var[T] &= E[T^2] - (E[T])^2 = E[(aX \pm bY)^2] - (aE[X] \pm bE[Y])^2 \\ &= E[a^2X^2 \pm 2abXY + b^2Y^2] - ((aE[X])^2 \pm 2abE[X]E[Y] + (bE[Y])^2) \\ &= a^2E[X^2] \pm 2abE[XY] + b^2E[Y^2] - a^2(E[X])^2 \mp 2abE[X]E[Y] - b^2(E[Y])^2 \\ &= a^2Var[X] + b^2Var[Y] \pm 2ab\Big(E[XY] - E[X]E[Y]\Big) \end{aligned}$$

- В процессе использовали свойство линейности математического ожидания.
- Слагаемое $E[XY] - E[X]E[Y]$ называется **ковариацией** X и Y.

Ковариация и независимость

Утверждение: Если X и Y - независимые случайные величины, то $E[XY] = E[X]E[Y]$

Используем определение независимости величин:

$$\forall x_i,\ \forall y_j:\ P_X\left(X = x_i\right) P_Y\left(Y = y_j\right) = P(\{X = x_i\} \cap \{Y = y_j\}) = P\left(X = x_i, Y = y_j\right)$$

$$\begin{aligned} E[XY] &= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j P(X = x_i, Y = y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j P_X(x_i) P_Y(y_j) \\ &= \sum_{i=1}^{n} x_i P_X(x_i) \sum_{j=1}^{m} y_j P_Y(y_j) = E[X]E[Y]. \end{aligned}$$

Будьте внимательны: обратное утверждение неверно!

Корреляция

- Ковариация не является хорошей мерой зависимости, поскольку она зависит от масштабов X и Y.
- Идея: нормализовать ковариацию и избавиться от масштабов.
- Корреляция:

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X]Var[Y]}}$$

- Она строго ограничена от -1 до 1 и:

$$Corr(X, Y) = 1 \leftrightarrow Y = kX + b,\ k > 0$$
$$Corr(X, Y) = -1 \leftrightarrow Y = -kX + b,\ k > 0$$

- Важно: Если переменные независимы, то их ковариация (и корреляция) равна нулю. Обратное утверждение НЕВЕРНО. Очень может быть, что коварициция равна нулю, но при этом величины являются зависимыми

Пример:

	$Y=3$	$Y=-3$
$X=-1$	$\frac{3}{4}$	0
$X=2$	0	$\frac{1}{4}$

1. Найдите маргинальные функции вероятности для X и Y.
2. Независимы ли X и Y?
3. Найдите ковариацию (и корреляцию) между X и Y.

Решение задачи

1. Маргинальные функции вероятности

Для X:

$$P(X=-1)=\sum_j P(X=-1,Y=y_j)=\frac{3}{4}+0=\frac{3}{4}$$

$$P(X=2)=\sum_j P(X=2,Y=y_j)=0+\frac{1}{4}=\frac{1}{4}$$

Для Y:

$$P(Y=3)=\sum_i P(X=x_i,Y=3)=\frac{3}{4}+0=\frac{3}{4}$$

$$P(Y=-3)=\sum_i P(X=x_i,Y=-3)=0+\frac{1}{4}=\frac{1}{4}$$

Пример

Проверка независимости

Проверим условие независимости для пары:

$$P(X=-1, Y=3) = \frac{3}{4} \quad \text{и} \quad P(X=-1)P(Y=3) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16} \neq \frac{3}{4}$$

Ответ: X и Y **не независимы**.

Пример

Ковариация и корреляция

Сначала найдем математические ожидания:

$$E[X] = (-1) \cdot \frac{3}{4} + 2 \cdot \frac{1}{4} = -\frac{3}{4} + \frac{1}{2} = -\frac{1}{4}$$

$$E[Y] = 3 \cdot \frac{3}{4} + (-3) \cdot \frac{1}{4} = \frac{9}{4} - \frac{3}{4} = \frac{3}{2}$$

Теперь найдем $E[XY]$:

$$\begin{aligned} E[XY] &= (-1) \cdot 3 \cdot \frac{3}{4} + (-1) \cdot (-3) \cdot 0 + 2 \cdot 3 \cdot 0 + 2 \cdot (-3) \cdot \frac{1}{4} \\ &= -\frac{9}{4} + 0 + 0 - \frac{6}{4} = -\frac{15}{4} \end{aligned}$$

Пример

Ковариация:

$$Cov(X,Y) = E[XY] - E[X]E[Y] = -\frac{15}{4} - \left(-\frac{1}{4}\right)\cdot\frac{3}{2}$$
$$= -\frac{15}{4} + \frac{3}{8} = -\frac{30}{8} + \frac{3}{8} = -\frac{27}{8}$$

Для корреляции найдем дисперсии:

$$E[X^2] = (-1)^2\cdot\frac{3}{4} + 2^2\cdot\frac{1}{4} = \frac{3}{4} + 1 = \frac{7}{4}$$
$$Var[X] = E[X^2] - (E[X])^2 = \frac{7}{4} - \left(-\frac{1}{4}\right)^2 = \frac{7}{4} - \frac{1}{16} = \frac{27}{16}$$

$$E[Y^2] = 3^2\cdot\frac{3}{4} + (-3)^2\cdot\frac{1}{4} = \frac{27}{4} + \frac{9}{4} = 9$$
$$Var[Y] = E[Y^2] - (E[Y])^2 = 9 - \left(\frac{3}{2}\right)^2 = 9 - \frac{9}{4} = \frac{27}{4}$$

Пример

Корреляция:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} = \frac{-\frac{27}{8}}{\sqrt{\frac{27}{16} \cdot \frac{27}{4}}}$$

$$= \frac{-\frac{27}{8}}{\sqrt{\frac{729}{64}}} = \frac{-\frac{27}{8}}{\frac{27}{8}} = -1$$

Корреляция равна -1, что означает полную отрицательную линейную зависимость между X и Y.