Связь случайных величин

Совместное распределение

Определение

Совместное распределение $X_1, \ldots, X_n$ (joint PMF)

— функция $PMF_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$, где $X_1, \ldots, X_n$ - дискретные случайные величины

Интерпретация

- совместное распределение показывает, как часто реализуется запрашиваемая комбинация значений случайных величин

Определение

Вероятность события в совместном распределении

— сумма вероятностей комбинаций значений случайных величин, для которых это событие выполнено

$$P(A) = \sum_{(x_1,\ldots,x_n)\in A} PMF_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

Определение

Маргинальное распределение

— распределение одной или нескольких случайных величин, полученное из совместного путём суммирования по значениям остальных величин.

Пример

$$P(X = x) = \sum_y P(X = x, Y = y)$$

Независимые события

Определение

Независимые события

События A и B называют **независимыми**, если $P(A \cap B) = P(A) \cdot P(B)$

Интерпретация

Это формализует наше интуитивное представление: если знание о том, что A произошло, не меняет вероятность B, то события независимы. Формально это выражается через условную вероятность:

$$P(B \mid A) = P(B)$$

А поскольку по определению условной вероятности:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

То подставляя, получаем:

$$\frac{P(A \cap B)}{P(A)} = P(B) \quad \Rightarrow \quad P(A \cap B) = P(A) \cdot P(B)$$

Определение

Независимые в совокупности события

События $A_1, .., A_n$ называют **независимыми в совокупности**, если **для любого подмножества** вероятность пересечения равна произведению вероятностей.

Пример

Попарная независимость ≠ независимость в совокупности
Бросают две монетки. Все исходы равновероятны.
События:

- F: первая — орёл → $\{HH, HT\}$
- S: вторая — орёл → $\{HH, TH\}$
- D : результаты разные → $\{HT, TH\}$
  Проверим попарную независимость:
- $P(F \cap S) = P(HH) = 0.25$, $P(F)P(S) = 0.5 \cdot 0.5 = 0.25$ ✅
- $P(F \cap D) = P(HT) = 0.25$, $P(F)P(D) = 0.5 \cdot 0.5 = 0.25$ ✅
- $P(S \cap D) = P(TH) = 0.25$, $P(S)P(D) = 0.5 \cdot 0.5 = 0.25$ ✅
  → События **попарно независимы**.
  Но:
- $P(F \cap S \cap D) = P(\varnothing) = 0$
- $P(F)P(S)P(D) = 0.5^3 = 0.125 \neq 0$
  ❌ Не выполняется условие полной независимости.
  Вывод: Попарная независимость **не влечёт** независимость в совокупности.

Независимость случайных величин

Определение

> **Независимые случайные величины**
>
> Дискретные случайные величины X,Y называют независимыми, если для любых их значений a и b независимы события $\{X = a\}$ и $\{Y = b\}$
>
> $$P(X = a, Y = b) = P(X = a) \cdot P(Y = b)$$

Определение

> **Независимые в совокупности случайные величины**
>
> Дискретные случайные величины $X_1, \dots X_n$ называются **независимыми в совокупности**, если каждые n-1 из них независимы в совокупности, и для любых их значений $a_1, \dots a_n$:
>
> $$P(X_1 = a_1, \dots, X_n = a_n) = P(X_1 = a_1) \cdot \dots \cdot P(X_n = a_n)$$

Пример

Дано совместное распределение:

Y\X	0	1	2
0	0.2	0	0.3
1	0.3	0.2	0

Найдём маргинальные вероятности:

- $P(X=0)=0.2+0.3=0.5$
- $P(Y=1)=0.3+0.2+0=0.5$
- $P(X=0,Y=1)=0.3$
- $P(X=0)\cdot P(Y=1)=0.5\cdot 0.5=0.25\neq 0.3$
  ⇒ Так как $P(X=0,Y=1)\neq P(X=0)\cdot P(Y=1)$, случайные величины **не являются независимыми**.

Следствие

Если случайные величины независимы, то по их маргинальным распределениям можно однозначно восстановить совместное распределение:
$P(X=x,Y=y)=P(X=x)\cdot P(Y=y)$

Для независимых в совокупности случайных величин любые события также независимы

Лемма

> **Независимость событий, порожденных независимыми случайными величинами**
>
> Если $X_1,\dots,X_n$ - независимые в совокупности дискретные случайные величины, а $S_1,\dots S_n\subseteq\mathbb{R}$, то события $\{X_1\in S_1\}$, $\{X_2\in S_2\}$, ... , $\{X_n\in S_n\}$ независимы в совокупности

Событие $\{X_k\in S_k\}$ надо понимать как "значение случайной величины X_k попало в множество S_k"

Пример

Если X_1 - число выпавших очков на кубике, а $S_1=\{1,3,5\}$, то $X_1\in S_1$- событие "выпало нечетное число"

Интерпретация

Если случайные величины **независимы в совокупности**, то **любые логические условия на их значения** (вроде "$X_1 > 5$", "$X_2 \in [0,1]$", "$X_3 = 0$") порождают **события, которые тоже независимы в совокупности**.
То есть вы можете:

- задать произвольные условия (множества SiSi),
- рассмотреть события {Xi∈Si}{Xi∈Si},
- и тогда **вероятность того, что все они произойдут одновременно**, будет равна **произведению вероятностей каждого события в отдельности**.

Лемма

> **Независимость функций от набор независимых случайных величин**
>
> Функции от непересекающихся наборов независимых в совокупности дискретных случайных величин задают независимые в совокупности случайные величины
> Если $X_1, .., X_n$ - независимые в совокупности случайные величины, а $g_1, g_2, .. g_k$ произвольные функции, то величины $g_1(X_1, \dots, X_{n_1})$, $g_2(X_{n_1+1}, \dots, X_{n_1+n_2}, \dots, g_k(X_{n-n_k+1}, \dots, X_n)$ будут независимы в совокупности

Определение

> **Матожидание в совместном распределении**
>
> Пусть $g(x_1, \dots, x_n)$ — функция, а $PMF_{X_1, \dots, X_n}$ — совместная PMF. Тогда:
>
> $$\mathbb{E}[g(X_1, \dots, X_n)] = \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) \cdot P(X_1 = x_1, \dots, X_n = x_n)$$

Лемма

> **Матожидание произведения независимых случайных величин**
>
> Для независимых случайных величин матожидание произведения равно произведению матожиданий

Ковариация и корреляция

Определение

Ковариация случайных величин X,Y

$$cov(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

Лемма

Свойства ковариации

- $$cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y, \quad \forall X, Y$$
- $$\text{X,Y независимы } \Rightarrow cov(X,Y) = 0$$
- $$cov(X,c) = 0, \quad c \in \mathbb{R}$$
- $$cov(X + c_1, Y + c_2) = cov(X,Y), \quad c_1, c_2 \in \mathbb{R}$$
- $$cov(X,X) = \mathbb{V}X$$
- $$cov(X,Y) = cov(Y,X)$$
- $$cov(c_1X, c_2Y) = c_1 \cdot c_2 \cdot cov(X,Y), \quad c_1, c_2 \in \mathbb{R}$$
- $$cov(aX + bY, cZ + dT) = ac \cdot cov(X,Z) + ad \cdot cov(X,T) + bc \cdot cov(Y,Z) + bd \cdot cov(Y,T)$$

Определение

Корреляция случайных величин X,Y

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Важно

Если одна из величин константа, то корреляция не определена. Однако обычно ее доопределяют нулем

Лемма

Область значений корреляции

$$corr(X, Y) \in [-1; 1], \quad \forall X, Y$$

Для независимых случайных величин попарные ковариации равны нулю, так что:

$$\begin{aligned} \mathbb{E}\big(a_1X_1 + \cdots + a_nX_n\big) &= a_1\mathbb{E}X_1 + \cdots + a_n\mathbb{E}X_n, \\ \mathrm{Var}\big(a_1X_1 + \cdots + a_nX_n\big) &= a_1^2\mathrm{Var}X_1 + \cdots + a_n^2\mathrm{Var}X_n, \\ \sigma_{a_1X_1+\cdots+a_nX_n} &= \sqrt{a_1^2\mathrm{Var}X_1 + \cdots + a_n^2\mathrm{Var}X_n}. \end{aligned}$$

Случайная выборка из распределения

Определение

Случайное семплирование из распределения размера n

— набор из n независимых в совокупности одинаково распределенных случайных величин

Определение

Случайная выборка из распределения размера n

— реализация случайного семплирования

Лемма

Сумма и среднее н.о.р.с.в

$$\mathbb{E}(X_1 + \dots + X_n) = n \cdot \mathbb{E}X;$$
$$\mathbb{V}(X_1 + \dots + X_n) = n \cdot \mathbb{V}X;$$
$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \mathbb{E}X;$$
$$\mathbb{V}(\overline{X}) = \mathbb{V}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\mathbb{V}X}{n};$$
$$\sigma_{X_1 + \dots + X_n} = \sqrt{n} \cdot \sigma_X;$$
$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}.$$

Flashcards

tags: #flashcardsSTAT

Совместное распределение

Что такое Совместное распределение $X_1, \dots, X_n$ (joint PMF)?
%
— функция $PMF_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$, где $X_1, \dots, X_n$ - дискретные случайные величины

Что такое Вероятность события в совместном распределении?
%
— сумма вероятностей комбинаций значений случайных величин, для которых это событие выполнено

$$P(A) = \sum_{(x_1,\dots,x_n)\in A} PMF_{X_1,\dots,X_n}(x_1, \dots, x_n)$$

Как найти $P(X = Y)$, зная совместное распределение?
%
Просуммировать $P(X = x, Y = x)$ по всем возможным x.

Как найти $P(X \leqslant Y)$ по совместному распределению?
%
Просуммировать все $P(X = x, Y = y)$, для которых $x \leq y$.

Что такое маргинальное распределение X?
%
$P(X = x) = \sum_y P(X = x, Y = y)$ — сумма по строке/столбцу таблицы.

Независимые события

Что такое Независимые события?
%
События A и B называют **независимыми**, если $P(A \cap B) = P(A) \cdot P(B)$

Что такое Независимые в совокупности события?
%
События $A_1, .., A_n$ называют **независимыми в совокупности**, если каждые n-1 из них независимы в совокупности, а также

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot \ldots \cdot P(A_n)$$

Независимость случайных величин

Что такое Независимые случайные величины?
%
Дискретные случайные величины X,Y называют независимыми, если для любых их значений a и b независимы события $\{X = a\}$ и $\{Y = b\}$

$$P(X = a, Y = b) = P(X = a) \cdot P(Y = b)$$

Что такое Независимые в совокупности случайные величины?
%
Дискретные случайные величины $X_1, \ldots X_n$ называются **независимыми в совокупности**, если каждые n-1 из них независимы в совокупности, и для любых их значений $a_1, \ldots a_n$:

$$P(X_1 = a_1, \ldots, X_n = a_n) = P(X_1 = a_1) \cdot \ldots \cdot P(X_n = a_n)$$

Сформулируйте лемму: Независимость событий, порожденных независимыми случайными величинами
%
Если $X_1, \ldots, X_n$ - независимые в совокупности дискретные случайные величины, а $S_1, \ldots S_n \subseteq \mathbb{R}$, то события $\{X_1 \in S_1\}$, $\{X_2 \in S_2\}$, ... , $\{X_n \in S_n\}$ независимы в совокупности

Сформулируйте лемму: Независимость функций от набор независимых случайных величин
%
Функции от непересекающихся наборов независимых в совокупности дискретных случайных величин задают независимые в совокупности случайные величины
Если $X_1, .., X_n$ - независимые в совокупности случайные величины, а $g_1, g_2, .. g_k$ произвольные функции, то величины $g_1(X_1, \ldots, X_{n_1})$, $g_2(X_{n_1+1}, \ldots, X_{n_1+n_2}, \ldots, g_k(X_{n-n_k+1}, \ldots, X_n)$ будут независимы в совокупности

Что такое Матожидание в совместном распределении?
%
Пусть $g(x_1, \ldots, x_n)$ — функция, а $PMF_{X_1, \ldots, X_n}$ — совместная PMF. Тогда:

$$\mathbb{E}[g(X_1, \ldots, X_n)] = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) \cdot P(X_1 = x_1, \ldots, X_n = x_n)$$

Сформулируйте лемму: Матожидание произведения независимых случайных величин
%
Для независимых случайных величин матожидание произведения равно произведению матожиданий

Ковариация и корреляция

Что такое Ковариация случайных величин X,Y?
%

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

Сформулируйте лемму: Свойства ковариации
%

- $$cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y, \quad \forall X, Y$$
- $$\text{X,Y независимы } \Rightarrow cov(X, Y) = 0$$
- $$cov(X, c) = 0, \quad c \in \mathbb{R}$$
- $$cov(X + c_1, Y + c_2) = cov(X, Y), \quad c_1, c_2 \in \mathbb{R}$$
- $$cov(X, X) = \mathbb{V}X$$
- $$cov(X, Y) = cov(Y, X)$$
- $$cov(c_1X, c_2Y) = c_1 \cdot c_2 \cdot cov(X, Y), \quad c_1, c_2 \in \mathbb{R}$$
- $$cov(aX + bY, cZ + dT) = ac \cdot cov(X, Z) + ad \cdot cov(X, T) + bc \cdot cov(Y, Z) + bd \cdot cov(Y, T)$$

Что такое Корреляция случайных величин X,Y?
%

$$corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Сформулируйте лемму: Область значений корреляции
%

$$corr(X, Y) \in [-1; 1], \quad \forall X, Y$$

Когда можно восстановить совместное распределение по маргинальным?
%
Когда X и Y независимы: $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$.

Если $\operatorname{cov}(X, Y) = 0$, следует ли, что X и Y независимы?
%
Нет. Обратное неверно. Только если независимы → ковариация = 0.

Может ли быть $\operatorname{corr}(X, Y) = 0$, но X и Y зависимы?
%
Да. Например, $Y = X^2$, X симметрично вокруг 0.

Чему равна $\operatorname{corr}(X, Y)$, если одна из величин — константа?
%
Не определена (деление на 0), но часто доопределяют как 0.

Как интерпретировать $\operatorname{corr}(X, Y) \approx -0.46$?
%
Умеренная отрицательная линейная связь: при росте одной величины другая в среднем снижается.

Чему равна $\operatorname{cov}(X, X)$?
%
$\operatorname{cov}(X, X) = \mathbb{V}X$

Случайная выборка из распределения

Что такое Случайное семплирование из распределения размера n?
%
— набор из n независимых в совокупности одинаково распределенных случайных величин

Что такое Случайная выборка из распределения размера n?
%
— реализация случайного семплирования

Сформулируйте лемму: Сумма и среднее н.о.р.с.в
%

$$\mathbb{E}(X_1 + \cdots + X_n) = n \cdot \mathbb{E}X;$$

$$\mathbb{V}(X_1 + \cdots + X_n) = n \cdot \mathbb{V}X;$$

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \mathbb{E}X;$$

$$\mathbb{V}(\overline{X}) = \mathbb{V}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\mathbb{V}X}{n};$$

$$\sigma_{X_1 + \cdots + X_n} = \sqrt{n} \cdot \sigma_X;$$

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}.$$

Чему равна дисперсия среднего $\bar{X}$ для н.о.р.с.в.?
%
$\mathbb{V}(\bar{X}) = \dfrac{\mathbb{V}X}{n}$

Почему дисперсия среднего уменьшается с ростом n?
%
Потому что усреднение "сглаживает" случайные отклонения.

Чем отличается "случайное семплирование" от "случайной выборки"?
%

- Семплирование — процесс (набор СВ),
- Выборка — реализация (набор чисел).

Как зависит $\sigma_{\bar{X}}$ от n?
%
$\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}}$ — уменьшается как $1/\sqrt{n}$.