Математика для Data Science. Теория вероятностей.
Условия задач

Содержание

Замечание. Таким цветом отмечены ссылки на сайт Stepik, а вот этим цветом — ссылки на страницы внутри этого файла.

Распределения и независимые случайные величины

Задача 1

Мы три раза подряд подбрасываем честную монету.

- Случайная величина Y – суммарное число выпавших решек
- Случайная величина Z – суммарное число выпавших орлов

1. Верно ли, что $Y = Z$?
2. Докажите, что $p_Y(a) = p_Z(a)$ для любого числа $a \in \mathbb{R}$. Тем самым, у Y и Z одинаковые функции вероятности. Можно записать это так: $p_Y = p_Z$.

Определение. Две случайные величины с совпадающими функциями вероятности называются *одинаково распределёнными*.

Задача 2

Дана случайная величина X. Пусть она принимает ровно n различных значений: $x_1, \ldots, x_n$.

1. Докажите, что $\forall i : p_X(x_i) \geq 0$
2. Докажите, что $p_X(a) = 0$ для всех a не равных одному из $x_1, \ldots, x_n$.
3. Докажите, что $\sum\limits_{i=1}^{n} p_X(x_i) = 1$. Или в другой записи: $p_X(x_1) + p_X(x_2) + \cdots + p_X(x_n) = 1$.

То есть про функцию вероятности можно думать, как про суммарную массу 1, как-то раскиданную по n точкам на числовой прямой.

Поэтому можно задавать функцию вероятности случайной величины, просто перечислив значения в точках $x_1, \ldots, x_n$. Например, такие условия: $p_Y(25) = 0.33, p_Y(38) = 0.67$ однозначно задают функцию вероятности случайной величины Y, которая принимает ровно два значения: 25 и 38.

Комментарий. Если эта задача вам кажется подозрительно простой, то да, подвоха нет, это действительно простая задача – фактически, мы просим вас применить определение функции вероятности случайной величины.

Независимые случайные величины

Задача 1

В конце прошлой недели мы поняли, что вычисление математического ожидания не перестановочно с умножением. То есть не всегда выполнено $E[X \cdot Y] = E[X] \cdot E[Y]$.

Задача. Докажите, что если X и Y независимы, то $E[X \cdot Y] = E[X] \cdot E[Y]$.

Дополнительная задача

Обозначение. Аналогично обозначению события $X = a$ введём следующее обозначение. Событие "случайная величина X приняла значение в отрезке $[a, b]$" будем обозначать так: $X \in [a, b]$.

Докажите, что если случайные величины X и Y независимы, то для любых $a, b, c, d \in \mathbb{R}$ события $X \in [a, b]$ и $Y \in [c, d]$ независимы.

Задача 2

Даны k совместно независимых случайных величин $X_1, \ldots, X_k$. Все эти величины имеют одинаковое распределение: $P(X_i = 0) = q, P(X_i = 1) = 1 - q$. Найдите распределение случайной величины $X_1 \cdot X_2 \cdot \ldots \cdot X_k$.

Дисперсия

Задача 2

1. Докажите, что для любой случайной величины X выполнено $Var[X] \geq 0$.
2. Докажите, что $Var[X] = 0$ если и только если X это постоянная случайная величина.

Задача 3

1. Пусть X и Y это независимые случайные величины. Докажите, что $Var[X + Y] = Var[X] + Var[Y]$.
2. Приведите пример X и Y, таких что $Var[X + Y] \neq Var[X] + Var[Y]$.

Задача 4

Пусть c это произвольное действительное число. Выразите дисперсию случайной величины cX через дисперсию случайной величины X.

Биномиальное распределение и стандартное отклонение

Задача 1

В вашей фирме работают n разработчиков. Каждый рабочий день каждый разработчик приходит в офис с вероятностью p, и с вероятностью $(1 - p)$ остаётся работать из дома. Разработчики приходят работать независимо друг от друга.

1. Какова вероятность, что сегодня конкретные k разработчиков будут работать из офиса, а все остальные – из дома? Например, если разработчиков зовут Аня, Гриша, Вика, Петя и Илья, то какова вероятность, что из разработчиков в офис придут только Аня и Петя?
2. Какова вероятность, что сегодня в офис придут ровно k разработчиков? Имена пришедших не уточняются – важно только чтобы пришло ровно k.
3. Найдите распределение случайной величины S, где S – число разработчиков, которые пришли сегодня в офис.

Задача 2

У вас та же фирма, что и в прошлой задаче. Пусть $X_i = 1$ если i-ый разработчик пришёл, и $X_i = 0$ в противном случае. Тогда $S := X_1 + \cdots + X_n$ это количество пришедших разработчиков.

1. Найдите $E[S]$.
2. Найдите $Var[S]$.

Задача 3

Теперь пусть разработчики либо все одновременно приходят (с вероятностью p), либо все одновременно не приходят (с вероятностью $1 - p$). Обозначим число пришедших разработчиков за T.

1. Найдите $E[T]$.
2. Найдите $Var[T]$.

Дополнительная задача

1. Зачем вообще искать средний квадрат отклонения? Давайте лучше искать само среднее отклонение, а не его квадрат. Найдите $E[X - E[X]]$.

2. Пусть мы приняли определение из прошлого шага $H[X] := E[|X - E[X]|]$. Мы хотим использовать H в наших вычислениях, а не дисперсию. Докажите, что даже для независимых X и Y не всегда выполнено $H[X + Y] = H[X] + H[Y]$.

Как мы помним, для независимых X и Y выполнено $Var[X + Y] = Var[X] + Var[Y]$. Так что дисперсия ещё и этим удобнее для вычислений, чем H (а не только тем, что H использует модуль, а Var нет).

Ряды

Задача 1

Докажите, что для любого $\beta \neq 1$ и любого $t \in \mathbb{N}$ выполнено $1 + \beta + \beta^2 + \cdots + \beta^{t-2} + \beta^{t-1} = \frac{1-\beta^t}{1-\beta}$.

Комментарий. Формулу из пункта 2 мы уже доказывали в курсе матана, когда говорили про градиентный спуск с моментом.

Задача 2

Дан ряд $1 + \beta + \beta^2 + \beta^3 + \cdots = \sum\limits_{n=0}^{\infty} \beta^n$. Для каждого β найдите сумму ряда или докажите, что ряд расходится.

Задача 3

1. Докажите, что ряд $\frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \ldots$ расходится.

2. Докажите, что ряд $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \cdots = \sum\limits_{n=1}^{\infty} \frac{1}{n}$ расходится.

Ряд из Пункта 2 называется *гармоническим*.

Комментарий. По аналогии с определениями пределами последовательностей можно сказать, что $\sum\limits_{n=1}^{\infty} \frac{1}{n} = +\infty$, ведь последовательность частичных сумм этого ряда сходится к $+\infty$. При этом такой ряд мы всё равно называем расходящимся.

Задача 4

Даны два сходящихся ряда: $\sum\limits_{n=1}^{\infty} a_n$ и $\sum\limits_{n=1}^{\infty} b_n$.

1. Докажите, что ряд $\sum\limits_{n=1}^{\infty} (a_n + b_n)$ сходится

2. Докажите, что ряд $\sum\limits_{n=1}^{\infty} (a_n - b_n)$ сходится

3. Докажите, что ряд $\sum\limits_{n=1}^{\infty} (ca_n)$ сходится для любого $c \in \mathbb{R}$

Естественно, в доказательстве можно использовать уже доказанные утверждения из курса матана.

Абсолютно сходящиеся ряды

Задача 1

Давайте потренируемся применять теорему с прошлого шага.
Пусть

- $a_1 = \frac{1}{1} - \frac{1}{2}$
- $a_2 = \frac{1}{3} - \frac{1}{4}$,
- $a_3 = \frac{1}{5} - \frac{1}{6}$,
- $a_4 = \frac{1}{7} - \frac{1}{8}$,
- и так далее

Ясно, что все $a_i \geq 0$. Давайте докажем, что ряд $\sum\limits_{n=1}^{\infty} a_n$ сходится.

Задача. Докажите, что последовательность частичных сумм ряда неубывающая и ограничена сверху числом 1.

Задача 2

В этой задаче мы докажем утверждения, ради которых и вводили теорему из пред-предыдущего шага. А эти утверждения помогут нам доказать некоторые свойства абсолютно сходящихся рядов.

Даны два ряда $\sum\limits_{n=1}^{\infty} a_n$ и $\sum\limits_{n=1}^{\infty} b_n$, при этом $0 \leq a_n \leq b_n$ для всех n. В таком случае говорят, что ряд $\sum\limits_{n=1}^{\infty} b_n$ *мажорирует* ряд $\sum\limits_{n=1}^{\infty} a_n$.

1. Докажите, что если $\sum\limits_{n=1}^{\infty} b_n$ сходится, то $\sum\limits_{n=1}^{\infty} a_n$ сходится. Приведите пример, когда обратное утверждение неверно.
2. Докажите, что если $\sum\limits_{n=1}^{\infty} a_n$ расходится, то $\sum\limits_{n=1}^{\infty} b_n$ расходится. Приведите пример, когда обратное утверждение неверно.

Задача 3

Докажем теорему:
Теорема. Любой абсолютно сходящийся ряд является сходящимся.

Пусть ряд $\sum\limits_{n=1}^{\infty} a_n$ абсолютно сходится, то есть ряд $\sum\limits_{n=1}^{\infty} |a_n|$ сходится. Давайте в три этапа докажем, что тогда $\sum\limits_{n=1}^{\infty} a_n$ сходится.

1. Докажите, что ряд $\sum\limits_{n=1}^{\infty} 2|a_n|$ неотрицателен и сходится.
2. Докажите, что ряд $\sum\limits_{n=1}^{\infty} (a_n + |a_n|)$ неотрицателен и мажорируется рядом $\sum\limits_{n=1}^{\infty} 2|a_n|$.
3. Докажите, что ряд $\sum\limits_{n=1}^{\infty} a_n = \sum\limits_{n=1}^{\infty} (a_n + |a_n|) - \sum\limits_{n=1}^{\infty} |a_n|$ сходится

Тем самым мы доказали, что из абсолютной сходимости ряда следует сходимость ряда.

Дополнительная задача

На этом шаге мы докажем теорему с пред-предыдущего шага:

Теорема. Если ряд абсолютно сходится к сумме S, то любой ряд, полученный из него перестановкой слагаемых, тоже абсолютно сходится к той же сумме S.

Пусть дан абсолютно сходящийся ряд $a_1+a_2+a_3+\cdots=\sum\limits_{n=1}^{\infty}a_n$. Обозначим его сумму за S, а его частичные суммы за $\{S_n\}$.

Обозначим за $b_1+b_2+b_3+\cdots=\sum\limits_{n=1}^{\infty}b_n$ ряд, полученный из ряда $\sum\limits_{n=1}^{\infty}a_n$ при помощи любой перестановки. Частичные суммы этого ряда будем обозначать за $\{\tilde{S}_n\}$.

В первых трёх пунктах давайте считать, что все $a_i \geq 0$. А в 4-ом и 5-ом пунктах мы поймём, что делать в случае, когда a_i могут быть отрицательными.

1. Пусть все $a_i \geq 0$. Докажите, что для любого i найдётся j, такой что $S_i \leq \tilde{S}_j$. Другими словами, для любой частичной суммы первого ряда найдётся не меньшая (то есть такая же или бо́льшая) частичная сумма второго ряда.

2. Пусть все $a_i \geq 0$. Докажите, что для любого i найдётся j, такой что $\tilde{S}_i \leq S_j$. Другими словами, для любой частичной суммы второго ряда найдётся не меньшая (то есть такая же или бо́льшая) частичная сумма первого ряда.

3. Пусть все $a_i \geq 0$. Докажите, что $\lim\limits_{n\to\infty}\tilde{S}_n$ существует и равен $S := \lim\limits_{n\to\infty}S_n$.

4. Теперь пусть a_i могут быть какими угодно. Докажите, что ряд, составленный только из неотрицательных членов ряда $\sum\limits_{n=1}^{\infty}a_n$ абсолютно сходится. Аналогично, докажите, что ряд, составленный только из отрицательных членов ряда $\sum\limits_{n=1}^{\infty}a_n$ абсолютно сходится.

5. Выведите из пункта 4, что $\lim\limits_{n\to\infty}\tilde{S}_n$ существует и равен $S := \lim\limits_{n\to\infty}S_n$.

Счётное пространство исходов

Задача 1

Мы бросаем честную монетку, пока не выпадет орёл. Обозначим орла за h и решку за t. У нас будут такие исходы и соответствующие вероятности:

1. $P(h)=\frac{1}{2}$
2. $P(th)=\frac{1}{4}$
3. $P(tth)=\frac{1}{8}$
4. $P(ttth)=\frac{1}{16}$
5. ...

Задача.

1. Найдите вероятность события ”число бросков больше 2.”

2. Найдите вероятность события ”число бросков делится на 3.”

Задача 2

В случае конечного количества исходов мы определяли пространства с равновероятными исходами.

Докажите, что не существует вероятностного пространства со счётным количеством исходов, такого что все исходы равновероятны.

Задача 3

Мы работаем с тем же вероятностным пространством, что и раньше. Мы подбрасываем монетку до первого орла и получаем такие исходы и соответствующие вероятности:

1. $P(h) = \frac{1}{2}$
2. $P(th) = \frac{1}{4}$
3. $P(tth) = \frac{1}{8}$
4. $P(ttth) = \frac{1}{16}$
5. $\ldots$

Задача. Определим случайную величину X так: $X(\underbrace{ttt\ldots t}_{k}h) = \frac{1}{3^{k+1}}$. Найдите $E[X]$.

Задача 4

Мы работаем в том же вероятностном пространстве, что и в предыдущей задаче. Определим ещё одну случайную величину Y так: $Y(\underbrace{ttt\ldots\ t}_{k}h) = (-1)^{k+1}$.

Найдите $E[Y]$.