

1 Основные понятия

1.1 Событие. Вероятность события

При рассмотрении опытов в которых могут быть различные исходы, результаты опытов будем называть событиями. Отличаем составные (разложимые) и элементарные (неразложимые) события.

Пример

Выпало 6 очков при броске двух игральных костей – составное разложение на (1, 5) или (2, 4) или (3, 3) или (4, 2) или (5, 1).

О том, что мы понимаем под элементарным событием надо предварительно условиться. Это неопределяемое понятие (как точка в геометрии). Они определяют идеализированный опыт. По определению, каждый неразложимый исход идеализированного опыта представляется одним и только одним элементарным событием. Совокупность всех элементарных событий называется пространством элементарных событий (σ). Элементарное событие – точки этого пространства. Событие – множество точек.

Совокупность точек представляет все те исходы, при которых происходит событие A , полностью описывают это событие. Любое множество точек A нашего пространства можно назвать событием; оно происходит или нет в зависимости от того, принадлежит или нет множеству A точка, представляющая исходный опыт.

Пример

Число курящих среди 100 человек. Пространство элементарных событий – множество чисел 0, 1, 2, . . . 100.

Определение

Невозможное событие – это событие, которое в результате данного опыта не может произойти. Обозначим его \emptyset . То есть, запись $A = \emptyset$ означает, что A не содержит элементарных событий.

Пример

Рассмотрим систему, состоящую из 6 атомов Н. Выбираем один атом.

Определение

Достоверное событие – то, которое в результате опыта обязательно произойдет. То есть $A = (\sigma)$.

Пример

Достоверное событие – выпадение ≤ 6 очков при броске одной игральной кости.

Определение

Событие, состоящее из всех точек, не содержащих событие A , называется собы-

тием противоположным A и обозначается \bar{A} .

$\bar{\sigma} = \emptyset$.

A – выпадение орла, \bar{A} – решки.

Определение

Суммой (объединением) $A + B$ ($A \cup B$) событий A и B назовем событие, которое состоит в том, что имеет место или A , или B , или (A и B). То есть это объединение множеств точек A и B .

Определение

Произведением (пересечением) $A \cdot B$ ($A \cap B$) событий A и B называется событие, которое состоит в том, что имеет место и A и B (одновременно) – пересечение множества точек A и B .



Определение

События A и B несовместны если $(A \cup B) = \emptyset$.

(То есть не могут произойти одновременно)

Пример

Бросок кости A – не менее 3 очков, B – не более 4 очков, \bar{A} – менее 3 очков (1 или 2), $A + B = \sigma$, $A \cdot B = 3$ или 4 очка.

Пример

Выстрел по мишени. A – попадание, B – промах, $A \cdot B = \emptyset$.

$$A + B = B + A$$

$$(A + B) + C = A + (B + C)$$

$$A \cdot B = B \cdot A$$

Определение

Пространство элементарных событий называется дискретным, если оно состоит из конечного числа точек, или из бесконечного числа точек, которые могут быть занумерованы последовательно (Счетное число точек).

Пример

Предыдущий пример – конечное число точек.

Теперь попробуем ввести вероятность, то есть число, которое характеризует степень объективной возможности события.

Определение

Пусть дано дискретное пространство элементарных событий σ с точками E_1 ,

$E_2, E_3 \dots$ Предполагаем, что с каждой точкой E_i (событием) связано число, называемое вероятностью E_i и обозначаемое $P(E_i)$, такое что:

$$1) P(E_i) \geq 0$$

$$2) P(E_1) + P(E_2) + \dots = 1$$

Вероятность любого события A есть сумма вероятностей элементарных событий из которых оно состоит.

$$1) P(\sigma) = 1$$

$$2) P(\emptyset) = 0$$

$$3) 0 \leq P(A) \leq 1$$

Как определить вероятность события в общей теории не постулируется. Об этом надо специально договариваться. Чаще всего встречается схема случаев.

Пусть пространство элементарных событий состоит из n точек, причем все они равновероятны, то есть по условиям симметрии есть основание считать, что ни одно из них не является объективно более возможным, чем другие. Напомним, кроме того, что элементарные события несовместны. Такие элементарные события обычно называют случаями.

Пример

Орел и решка при броске монеты. Появляется для любой из карт тщательно перетасованной колоды.

Пусть событие состоит из m точек (эти m случаев называются благоприятными событию A). Тогда вероятность $P(A) = \frac{m}{n}$

Пример

Бросок игральной кости. A – выпадение четного числа очков.

$$n = 6, m = 3 (2, 4, 6), \text{ следовательно, } P(A) = \frac{3}{6} = \frac{1}{2}$$

В других ситуациях, не сводящихся к схеме случаев, вероятность определяется по другому (например плотник, землемер, штурман измеряют расстояния – одно и тоже, но делают это по разному). При этом все способы с корнями уходят в опыт.

Пусть производится n опытов, в каждом из которых может появиться событие. Частотой события называется отношение числа опытов, в которых появилось к общему числу опытов. Частоту часто называют статистической вероятностью.

$$0 \leq P^*(A) \leq 1, P^*(A) = \frac{m}{n}$$

Так определенная статистическая вероятность носит случайный характер. Но при росте n она стабильно около некоторого значения. При $n \rightarrow \infty$ с практической достоверностью (то есть, вероятность ошибки сколь угодно мала) можно утверждать, что частота события будет сколько угодно мало отличаться от вероятности его в отдельном опыте. Более подробно это рассмотрим потом.

Факты из комбинаторики

Число размещений с повторениями. $\overline{A}_n^k = n^k$, где n – количество типов элементов – группы по k элементов с учетом порядка (число трехзначных чисел в десятичной системе счисления равно $10^3 - 10^2$).

Число размещений без повторений $A_n^k = \frac{n!}{(n-k)!}$

Пример

12 человек участвует в соревновании. Сколько вариантов распределения медалей. $\overline{A}_{12}^3 = 12 \cdot 11 \cdot 10 = 1320$

Число сочетаний без повторений $C_n^k = \frac{n!}{(n-k)!(k)!}$

Пример

12 команд. Сколько способов сформировать финальную группу из 3 команд без учета мест?

$$C_{12}^3 = \frac{12 \cdot 11 \cdot 10}{2 \cdot 3} = 220$$

Число перестановок из n различных элементов $P_n = n! = A_n^n$

Число перестановок из $n = n_1 + n_2 + n_3 + \dots + n_k$ элементов, $n_1 - 1$ типа, $n_2 - 2$ типа, \dots $n_k - k$ -го типа,

$$P(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}$$

Пример

8 ладей расставляют на доске. Какова вероятность, что никакие две не бьют друг – друга?

Сколько способов расположить ладей на шахматной доске что бы они не били друг друга? На каждой горизонтальной по одной. Пусть на первой горизонтальной она стоит на позиции a_1 , на второй на позиции a_2 и так далее.

(a_1, a_2, \dots, a_8) – перестановка чисел $1, 2, \dots, 8$

То есть благоприятных случаев $P_8 = 8!$

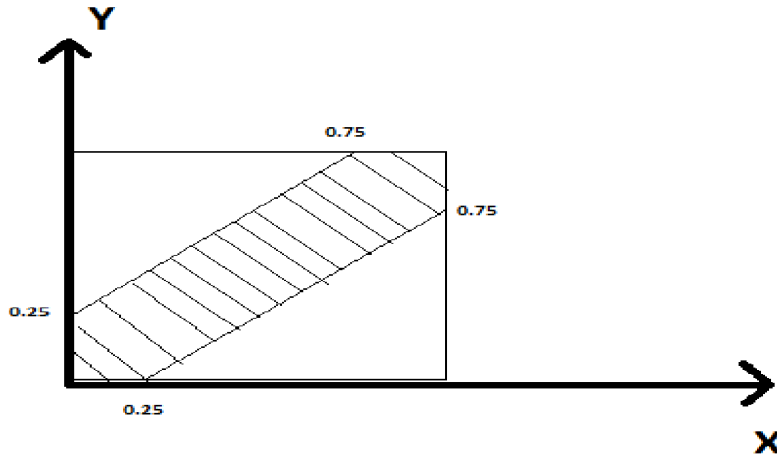
$$P = \frac{8!}{8^2 \cdot (8^2 - 1) \cdot (8^2 - 2) \cdot \dots \cdot (8^2 - 7)} \approx 9 \cdot 10^{-6}$$

Как определить вероятность если пространство элементарных событий не является конечным? Часто здесь имеет смысл метод геометрической вероятности. Если пространство σ может быть изображено геометрической фигуры и по условию опыта вероятность попадания точки (элементарного события) в любую часть области σ пропорционально мере этой части (длине, площади, объему \dots) и не зависит от ее расположения и формы, то вероятность события A определяется как $P(A) = \frac{S_A}{S}$, где S_A – мера части области, попадание в которую благоприятствует событию, S – мера всей области.

Пример

Двое договорились встретиться в определенном месте между 17 и 18 часами. Пришедший первым ждет второго 15 минут, после чего уходит. Определить вероятность встречи, если время прихода каждого независимо и равномерно в

течении этого часа.



Благоприятные исходы : $|x - y| \leq \frac{1}{4}$

$$\frac{1}{4} \leq x - y \leq \frac{1}{4}$$

$$S = 1 - \left(1 - \frac{1}{4}\right)^2$$

$$P = \frac{1 - \frac{9}{16}}{1} = \frac{7}{16}$$

Парадокс де-Мере-Паскаля

Что вероятнее: при 3 бросках игральной кости получить в сумме: 11 или 12 очков?

Рассуждение де-Мере: Суммы 11 и 12 образуются при выпадении на костях следующих цифр: $12 = 6 + 5 + 1 = 6 + 3 + 3 = 5 + 4 + 3 = 5 + 5 + 2 =$

$4 + 4 + 4$ (то есть 6 вариантов);

$11 = 6 + 4 + 1 = 6 + 3 + 2 = 5 + 5 + 1 = 5 + 4 + 2 = 5 + 3 + 3 =$

$4 + 4 + 2$ (то есть 6 вариантов);

То есть 11 и 12 должны быть равновероятны, но на опыте 11 появляется чаще.

На ошибку указал Паскаль: необходимо учитывать все возможные комбинации цифр, дающие в сумме 11 или 12.

Например $6 + 5 + 1 = 6 + 1 + 5 = 5 + 1 + 6 = 5 + 6 + 1 = 1 + 5 + 6 = 1 + 6 + 5$ (то есть 6 способов = $3!$). Аналогично, $6 + 4 + 2$ (6 = $3!$ способов), $6 + 3 + 3$ (3 способа), $5 + 4 + 3$ (6 способов), $4 + 4 + 4$ (1 способ).

$$P_{11} = \frac{6+6+6+3+3+3}{6^3} > P_{12}.$$

Сравнение статистик Больцмана, Бозе – Энштейна, Ферми – Дирака

Дано k частиц и l ячеек ($l > k$)

Найти вероятность того что:

1) в определенных k ячейках окажется по 1 частице

2) в каких то ячейках окажется по одной частице

Статистика Больцмана

Ей подчиняется обычный газ.

Условия:

а) частицы различны

б) в любой ячейке может находиться сколько угодно частиц

1) Общее число исходов l^k (Так как любую частицу можно положить в любую ячейку). Благоприятных исходов $k!$, так как частицы различны и их можно переставлять.

Значит $P_1 = \frac{k!}{l^k}$

2) Теперь можно k ячеек выбирать из l . (То есть число сочетаний k из общего числа l). То есть число благоприятных исходов равно $C_l^k \cdot k!$

$P_2 = \frac{C_l^k \cdot k!}{l^k}$

Статистика Бозе-Энштейна

Ей подчиняется обычный газ.

Условия:

а) частицы различны

б) в любой ячейке может находиться сколько угодно частиц

Общее число исходов.

Переставив ячейки в ряд, границы определим перегородками, которых $l + 1$. Если поменять местами две частицы, то нового распределения не получится, так как частицы неразличимы. Если поменять местами две перегородки, то тоже ничего нового не получится, так как все перегородки одинаковы. Если же поменять местами перегородку и частицу, то получится новое распределение. Две крайние перегородки закреплены, поэтому в перестановке участвует $l - 1$ перегородок и k частиц, то есть $k + l - 1$ элементов.

Число перестановок равно : $\frac{(k+l-1)!}{k! \cdot (l-1)!}$

Число благоприятных исходов равно:

1) Так как перестановка частиц не дает нового распределения, то благоприятный исход один (при фиксированных ячейках).

То есть $P_1 = \frac{1}{\frac{(k+l-1)!}{k! \cdot (l-1)!}} = \frac{k! \cdot (l-1)!}{(k+l-1)!}$

2) Число благоприятных исходов равно числу способов выбрать k ячеек из l , где будут частицы: $C_l^k = \frac{l!}{k! \cdot (l-k)!}$

$$P_2 = \frac{l! \cdot (k+l-1)!}{k! \cdot (l-k)! \cdot k! \cdot (l-1)!}$$

Статистика Ферми–Дирака

Ей подчинен, например, электронный газ.

а) частицы неразличимы

б) в ячейке может находиться не более одной частицы (принцип Паули).

Общее число исходов – это число способов выбрать из l ячеек k , где будут частицы, то есть $C_l^k = \frac{l!}{k! \cdot (l-k)!}$

Число благополучных исходов

1) Так как k ячеек определены, а частицы неразличимы, то благополучный исход один.

$$P_1 = \frac{1}{C_l^k} = \frac{k! \cdot (l-k)!}{l!}$$

2) Число благополучных исходов равно числу способов выбрать k заполненных ячеек из l равно C_l^k , следовательно $P_2 = \frac{C_l^k}{C_l^k} = 1$

1.2 Основные теоремы теории вероятностей

Теорема сложения вероятностей

$$P(A + B) = P(A) + P(B) - P(AB)$$

Доказательство: (для дискретного пространства элементарных событий). Что – бы получить $P(A + B)$ надо сложить вероятности точек входящих в A , и точек, входящих в B , на каждую по одному разу. Точки из AB сосчитали дважды, поэтому их надо вычесть.

Если A и B несовместны то $P(A + B) = P(A) + P(B)$. Если A_1, A_2, \dots, A_n – попарно несовместны, то $P(\sum_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

Определение:

Говорят, что события A_1, A_2, \dots, A_n образуют полную группу, если $A_1 + A_2 + \dots + A_n = \sigma$.

Следствие 1

Если события A_1, A_2, \dots, A_n образуют полную группу попарно несовместных событий, то $\sum_{i=1}^n P(A_i) = 1$

Следствие 2

Сумма вероятностей противоположных событий равна 1: $P(A) + P(\bar{A}) = 1$

Если событий три:

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i,j} P(A_i A_j) + \dots + (-1)^{n-1} P(A_1, A_2, \dots, A_n)$$

Пример

Есть 100 карточек с числами 1, 2, 3, . . . 100. Случайно выбирается одно из них. Событие A – число делится на 2, B – делится на 3. Найти вероятность $P(A+B)$.

$$P(A) = \frac{50}{100}, P(B) = \frac{33}{100}.$$

$$AB - \text{делится на 6: } P(AB) = \frac{16}{100}.$$

$$P(A+B) = \frac{50}{100} + \frac{33}{100} - \frac{16}{100} = \frac{67}{100}.$$

Теорема умножения вероятностей

Определение

Событие A называется независимым от события B , если вероятность события A не зависит от того, произошло событие B или нет, и зависимым, если вероятности A меняются в зависимости от того, произошло событие B или нет.

Пример 1

Бросание 2 монет. A – орел на первой монете, B – орел на второй монете. Эти события независимы.

Пример 2

Охотник, имеющий один патрон попадает в цель (событие A). Событие B – лев ловит добычу – зависимые, если добыча – охотник.

Определение

Вероятность события A , вычисленная при условии, что имело место событие B – называется условной вероятностью события A , и обозначается $P(A|B)$, или $P_B(A)$ (считается $P(B) \neq 0$). Условие независимости события A от B это $P(A|B) = P(A)$.

Теорема

$$P(A \cdot B) = P(A) \cdot P(B|A).$$

Доказательство:

Для схемы случаев. Пусть в результате опыта возможно n исходов, событие A благоприятных m исходов, событие B – k исходов, и событие A , и событие B

– l исходов. $P(A \cdot B) = \frac{l}{n}$, $P(A) = \frac{m}{n}$. Если известно, что A происходит, то из n остальных возможных исходов, из них l блокируют событие B , то есть $P(A|B) = \frac{l}{m}$. То есть $P(A \cdot B) = P(A) \cdot P(B)$.

Ясно, что $P(A \cdot B) = P(B) \cdot P(A|B)$

Следствие 1

Если событие A не зависит от события B , то и событие B не зависит от события A .

Известно, что $P(A) = P(A|B)$. Считаем, что $P(A) \neq 0$.

$$\begin{cases} P(AB) = P(A)P(B|A) \\ P(AB) = P(B)P(A|B) \end{cases}$$

Следовательно, $P(A)P(B|A) = P(B)P(A)$, следовательно

(при условии $P(A) \neq 0$) $P(B|A) = P(B)$

То есть, свойство зависимости или независимости взаимно.

Следствие 2

Для независимых событий $P(A \cdot B) = P(A) \cdot P(B)$

Замечание: Эта формула может быть взята за определение независимых событий.

Умножение для n событий.

$$P(A_1, A_2, \dots, A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 A_2) \dots P(A_n|A_1, A_2, \dots, A_{n-1})$$

Определение

События A_1, A_2, \dots, A_n называются независимыми в совокупности, если :

$$P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n)$$

Замечание: это определение эквивалентно следующему: события независимы в совокупности, если любое из них не зависит от любой совокупности остальных.

Независимость в совокупности не эквивалентна попарной независимости событий

Пример 1 (Пример Берштейна):

Имеется 4 шара. Красного, желтого, зеленого и трехцветный, имеющий красный, жёлтый и зелёный цвета на себе. Событие K - вынули шар, на котором красный цвет, событие $Ж$ - есть желтый, событие $З$ - зелёный. $P(K) = \frac{1}{2} = P(Ж) = P(З)$

Вероятность того, что вынули шар одновременно с двумя цветами $= P(K \cdot Ж) = \frac{1}{4} = P(K) \cdot P(Ж)$, $P(K \cdot З) = P(K) \cdot P(З) = \frac{1}{4}$, $P(Ж \cdot З) = P(Ж) \cdot P(З) = \frac{1}{4}$. Вероятность того, что выпадет три цвета равна $P(K \cdot Ж \cdot З) = \frac{1}{4} \neq$

$\neq P(K) \cdot P(Ж) \cdot P(З) = \frac{1}{8}$ То есть они попарно независимы, но зависимы в

совокупности.

Пример 2

Техническое устройство отказывает с вероятностью $p = 0.5$. Сколько раз его надо продублировать, что бы вероятность отказа установки была $q < 0.1$?

$$P = p^n < 0.1 = q, n > \frac{\ln(q)}{\ln(p)}$$

Суммы и произведение вероятностей часто работают вместе.

Пример 3

Пусть все элементы отказывают независимо от комбинаций других. Чему равна вероятность отказа цепи?

$$P(A + B_1B_2 + C_1C_2C_3) = P(A) + P(B_1) \cdot P(B_2) + P(C_1C_2C_3) - P(AB_1B_2) - P(AC_1C_2C_3) - P(B_1B_2C_1C_2C_3) + P(AB_1B_2C_1C_2C_3)$$

Формула полной вероятности

Определение

Говорят, что события H_1, H_2, \dots, H_n образуют полную группу, если $H_1 + H_2 + \dots + H_n$ – достоверное событие.

Пусть H_1, H_2, \dots, H_n – полная группа попарно несовместимых событий. Эти события будем называть гипотезами. Пусть надо найти вероятность события A , которое может произойти вместе с одной из гипотез. Тогда $P(A) = \sum_{i=1}^n P(H_i) \cdot$

$$P(A|H_i)$$

Так как H_1, \dots, H_n – полная группа. $A = H_1A + H_2A + \dots + H_nA$. H_1, \dots, H_n – попарно несовместные, следовательно H_1A, \dots, H_nA – несовместны, следовательно $P(A) = P(H_1A) + \dots + P(H_nA) = \sum_{i=1}^n P(H_iA)$, следовательно, по теореме

$$\text{умножения } P(A) = \sum_{i=1}^n P(H_i) \cdot P(A|H_i)$$

Пример

По самолету производится 3 выстрела. Вероятность попадания при первом 0.4, при втором – 0.5, при третьем – 0.7. Для вывода самолета из строя заведомо достаточно 3 попаданий. При первом попадании самолет выходит из строя с вероятностью 0.2, при двух – 0.6. Найти вероятность того, что в результате трех выстрелов самолет будет выведен из строя.

H_i – в самолет попал i -й снаряд, $P(H_0) = 0.6 \cdot 0.5 \cdot 0.3 = 0.09$. $P(H_1) = 0.4 \cdot 0.5 \cdot 0.3 + 0.6 \cdot 0.5 \cdot 0.3 + 0.6 \cdot 0.5 \cdot 0.7 = 0.36$

$P(H_2) = 0.6 \cdot 0.5 \cdot 0.7 + 0.4 \cdot 0.5 \cdot 0.7 + 0.4 \cdot 0.5 \cdot 0.5 = 0.41$, $P(H_3) = 0.4 \cdot 0.5 \cdot 0.7 = 0.14$

$P(A) = 0.36 \cdot 0.2 + 0.41 \cdot 0.6 + 0.14 \cdot 1 = 0.458$

Формула Байеса

Пусть событие A может произойти с одним из n попарно несовместимых событий H_1, H_2, \dots, H_n , образующих полную группу. Вероятности гипотез до опыта известны – $P(H_1), \dots, P(H_n)$ (априорные вероятности). Произведен опыт, в результате которого произошло событие A . Как следует изменить вероятности гипотез в связи с появлением A (то есть найти постериорные вероятности $P(H_i|A)$)? $P(AH_i) = P(A) \cdot P(H_i|A) = P(H_i) \cdot P(A|H_i)$, следовательно $P(H_i|A) = \frac{P(H_i) \cdot P(A|H_i)}{P(A)}$.

$$\text{Итого: } P(H_i|A) = \frac{P(H_i) \cdot P(A|H_i)}{\sum_{j=1}^n P(H_j) \cdot P(A|H_j)}$$

Пример

В первой урне 5 белых и 10 черных шаров. Во второй - 3 белых и 7 черных. Из второй в первую переложили 1 шар, а затем из второй вынули один шар. Оказалось что он белый. Найти вероятность того, что был переложён белый шар H_1 - переложили белый, H_2 - черный, $P(H_1) = \frac{3}{10}$, $P(H_2) = \frac{7}{10}$. $P(A|H_1) = \frac{6}{11}$, $P(A|H_2) = \frac{5}{11}$

$$P(A) = \frac{3}{10} \cdot \frac{6}{11} + \frac{7}{10} \cdot \frac{5}{11}, P(H_1|A) = \frac{\frac{3}{10} \cdot \frac{6}{11}}{\frac{3}{10} \cdot \frac{6}{11} + \frac{7}{10} \cdot \frac{5}{11}} = \frac{18}{18+35}$$

Пример

Известно, что 5% мужчин и 0.25% женщин - дальтоники. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность того, что это мужчина?

$$P(H_1) = P(H_2) = \frac{1}{2}$$

$$P(A|H_1) = 0.05, P(A|H_2) = 0.0025, P(H_1|A) = \frac{\frac{1}{2} \cdot 0.05}{\frac{1}{2} \cdot 0.05 + \frac{1}{2} \cdot 0.0025} = \frac{20}{21}$$

1.3 Повторение опытов

В самом начале развития теории вероятностей выяснилась фундаментальная роль одной математической схемы, изученной швейцарцем Яковом Бернулли. Схема такая: проведем последовательность испытаний, в каждом из которых вероятность события A одна и та же (p). Испытания независимы, то есть вероятность выявления события A в каждом из них не зависит от того, появилось оно или нет в других испытаниях.

Пример

2 игрока играют в шахматы 3 партии. Вероятность выигрыша первого $p = \frac{2}{3}$. Найти вероятность того, что он выиграет 2 партии.

$$\text{Это можно осуществить : } P(A_1 A_2 \bar{A}_3 + \bar{A}_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}.$$

В общем виде: вероятность того, что в n опытах событие произойдет m раз $P_n(m)$ равна $A_1 A_2 \dots A_m \bar{A}_{m+1} \dots \bar{A}_n + A_1 A_2 \dots \bar{A}_m A_{m+1} \bar{A}_{m+2} \dots \bar{A}_n + \dots + \bar{A}_1 A_2 \dots \bar{A}_{n-m} A_{n-m+1} \dots A_n$ В каждую комбинацию A входит m раз Число комбинаций C_n^m , все комбинации несовместны, следовательно $P_n(m) = p^m q^{n-m} + p^m q^{n-m} + \dots = C_n^m p^m q^{n-m}$, где $q = 1 - p$.

Так как по форме $P_n(m)$ член разложения бинома $(q + p)^n$ распределение вероятностей такого вида называется биномиальным распределением.

Пример

Два шахматных игрока, 10 результативных партий (ничьи не учитываются), Вероятность выигрыша первого - $\frac{2}{3}$, второго - $\frac{1}{3}$. Найти вероятность выигрыша

всей игры первым?

$$P_1 = P_{10}(6) + P_{10}(7) + P_{10}(8) + P_{10}(9) + P_{10}(10) = \frac{2^6}{3^{10}} \cdot (210 + 240 + 180 + 80 + 16) = \frac{2^6 \cdot 241}{3^9}$$

$$P_2 = P_{10}(6) + P_{10}(0) + P_{10}(1) + P_{10}(2) + P_{10}(3) + P_{10}(4) = \frac{1507}{3^9} \cdot (210 + 240 + 180 + 80 + 16)$$

То есть вероятность выигрыша первой партии у первого в два раза больше чем у второго, вероятность выигрыша матча у первого в 10 раз больше, чем у второго.

Замечание

$$\sum_{m=0}^n P_n(m) = 1, (p+q)^n = 1, p+q = 1, \text{ следовательно, } (p+q)^n = \sum_{m=0}^n P_n(m) =$$

$$\sum_{m=0}^n C_n^m p^m q^{(n-m)} - \text{Бином Ньютона.}$$

2 Случайные величины

Определение

Функция, определённая на пространстве элементарных событий называется случайной величиной.

Определение

Случайная величина называется дискретной, если она определена на дискретном пространстве элементарных событий.

Замечание: лучше бы было называть случайные величины функциями случая

Пример

Дискретная случайная величина. Число тузов у одного игрока при игре в бридж, число совпадающих дней рождения в группе из n человек.

Обозначать случайные величины будем буквами X, Y, \dots и их значения x, y, \dots . Пусть X – случайная величина. x_1, x_2, \dots – ее значения. Совокупность всех элементарных событий, на которых X принимает значение x_i образует событие $X = x_i$. Его вероятность обозначается $P(X = x_j) = p_j$. Соотношение, устанавливающее связь между значениями случайных величин и их вероятностями называется законом распределения случайной величины. Самой простой формой закона распределения для случайной величины является ряд распределения то есть таблица распределения, в которой сведены значения случайной величины и их вероятности. Для наглядности это часто изображают на графике и точки соединяют отрезками прямых. Получившаяся фигура называется многоугольником распределения. Так как события $X = x_j$ – несовместимы и образуют полную группу, то $\sum_{i=1}^n p_i = 1$.

Пример

Баскетболист бросает мяч в кольцо до первого попадания, либо пока не сделано

3 броска. Вероятность попадания при одном броске равна 0.7.

Функция распределения

Вероятность $P(X = x)$ часто использовать неудобно. Используют вероятность $P(X < x)$.

Определение

Функция распределения (интегральная функция распределения) или интегральный закон распределения или распределение накопленной вероятности - это функция на вещественной оси, определяемая следующим образом: $F(x) = P(X < x)$. $F(x)$ полностью характеризует случайную величину с вероятностной точки зрения, то есть это форма закона распределения.

Свойства:

- 1) $F(x)$ - неубывающая функция (то есть, при $x_1 < x_2$ $F(x_1) \leq F(x_2)$)
- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$
- 3) $\lim_{x \rightarrow +\infty} F(x) = 1$
- 4) $F(x)$ непрерывна слева. $F(x_i) = \lim_{x \rightarrow x_i - 0} F(x)$

Определение

Случайная величина называется непрерывной, если ее функция распределения непрерывна.

Для дискретной случайной величины: $F(x) = \sum_{x_j < x} P(X = x_j)$

Пример

Случайная величина – площадь разрушений, наносимых бомбой. Значение этой случайной величины непрерывно заполняет промежуток от 0 до πR^2 (R – радиус действия). Но в точках 0 и πR^2 у функции распределения скачки, так как этим значениям соответствуют конечные вероятности (вероятности положений 1 и 3 соответственно) круга разрыва.

Вероятность попадания точки на отрезок.

$$P(\alpha \leq x < \beta) = P(x < \beta) - P(x < \alpha) = F(\beta) - F(\alpha)$$

Замечание

$$P(\alpha \leq x \leq \beta) = F(\beta) - F(\alpha) + P(x = \beta)$$

$$P(x = \alpha) = \lim_{\beta \rightarrow \alpha} P(\alpha \leq x < \beta) = \lim_{\beta \rightarrow \alpha} (F(\beta) - F(\alpha))$$

Если слева величина непрерывна, то $P(x = \alpha) = 0$.

Плотность распределения

Пусть непрерывная случайная величина имеет дифференцируемую функцию распределения. Тогда $P(x < X < x + \Delta x) = F(x + \Delta x) - F(x)$

Определение

$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = F'(x)$ – плотность вероятности (плотность распределения).

$F(x)$ – первообразная для $f(x)$. Так как $F(-\inf) = 0$, то константа в первообразной выбирается однозначно.

$$F(x) = \int_{-\infty}^x f(x) dx$$

Найдем $P(\alpha < X < \beta)$ через $f(x)$. Так как вероятность одного значения непрерывной случайной величины равна нулю: $P(\alpha < X < \beta) = P(\alpha \leq X < \beta) =$

$$= F(\beta) - F(\alpha) = \int_{-\infty}^{\beta} f(x) dx - \int_{-\infty}^{\alpha} f(x) dx = \int_{\alpha}^{\beta} f(x) dx. \text{ Геометрически } P(\alpha < x < \beta) \text{ -- это площадь.}$$

Свойства плотности вероятности:

1) $f(x) \geq 0$ (следует из того, что $F(x)$ неубывает

2) $\int_{-\infty}^{\infty} f(x) dx = 1$ (так как $F(\infty)$)

То есть график плотности вероятности выше оси абсцисс, а полная площадь равна 1.

Пример

$$f(x) = \begin{cases} a \cdot \cos(x), & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 0, & x < -\frac{\pi}{2} \text{ или } x > \frac{\pi}{2} \end{cases}$$

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a \cdot \cos x dx = 2a = 1, \text{ следовательно, } a = \frac{1}{2}$$

$$F(x) = \begin{cases} 0, & x < -\frac{\pi}{2} \\ \frac{1}{2} \cdot (\sin(x) + 1), & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 1, & x > \frac{\pi}{2} \end{cases}$$

Числовые характеристики случайных величин.

Функция распределения или плотность полностью описывает случайную величину с вероятностной точки зрения.

Характеристика	Дискретная сл. вел-на X	Непрерывная сл. вел-на X
Мат ожидание	$\sum_k x_k p_k$	$\int_{-\infty}^{\infty} x f(x) dx$
Дисперсия	$\sum_k (x_k - M(x))^2 p_k$	$\int_{-\infty}^{\infty} (x - M(x))^2 f(x) dx$
Начальный момент порядка S	$\sum_k x_k^s p_k$	$\int_{-\infty}^{\infty} x^s f(x) dx$
Центральный момент порядка S	$\sum_k (x_k - M(x))^s p_k$	$\int_{-\infty}^{\infty} (x - M(x))^s f(x) dx$
Мода	наиболее вероятное значение	значение, в котором f(x) максимально
Медиана	не определено	такое $X = Me$, что $P(X < Me) =$ $= P(X > Me)$
Среднеквадратичное отклонение $\sigma(x)$	$\sigma(x) = \sqrt{D(x)}$	$\sigma(x) = \sqrt{D(x)}$
Коэффициент симметрии S_k	$S_k = \frac{\mu_3}{\sigma^3}$	$S_k = \frac{\mu_3}{\sigma^3}$
Эксцесс E_x	$E_x = \frac{\mu_4}{\sigma^4}$	$E_x = \frac{\mu_4}{\sigma^4} - 3$

Определение

Математическое ожидание дискретной случайной величины (среднее значение) – это число $M(X)$ ($E(X)$, \bar{X} , $< X >$), определяется по формуле $M(X) = \sum_k x_k p_k$, при условии, что ряд абсолютно сходится. Если ряд абсолютно расходится, то говорят что X не имеет конечного математического ожидания. Пусть случайная величина может принимать одно из N равновероятных

значений x_1, x_2, \dots, x_n . Тогда $p_i = \frac{1}{N}$ и $M(X) = \frac{\sum_{i=1}^N x_i}{N}$ – среднее арифметическое. Если значение не является равновероятным, то надо брать взвешенное среднее, что и сделано в определении. Аналогично с точечными массами и центром масс.

Замечание

$M(X)$ может быть значением случайной величины.

Пример

Испытываются однотипные приборы. Вероятность каждого пройти испытание равно p и независимы. Испытания заканчиваются после выхода из строя перво-

го же прибора. X – число произведенных испытаний. Чему равно $M(X)$?

$$P(X = k) = q \cdot p^{k-1}, \quad q = 1 - p.$$

$$M(X) = 1 \cdot q + 2q \cdot p + 3q \cdot p^2 + \dots + k \cdot q \cdot p^{k-1} = q(1 + 2p + 3p^2 + \dots)' = q \cdot \left(\frac{p}{1-p}\right)' = \frac{q}{(1-p)^2} = \frac{1}{q}$$

Пример

Пример отсутствия конечного математического ожидания – "Петербургская игра". Бросается монета до тех пор, пока не выпадет орёл. Если это случается при бросании с номером r , то игрок получит 2^r рублей. $x_r = 2^r$, $p_r = 2^{-r}$.

$$\sum_{r=1}^{\infty} x_r p_r = \sum_{r=1}^{\infty} 1 = \infty$$

Определение

Мода – координата максимума $f(x)$. Если кривая распределения имеет более одного максимума, то распределение называется полимодальным.

Определение

Медиана : $P(X < Me) = P(X > Me)$, для непрерывной случайной величины.

Определение

Квантиль порядка p – это значение x_p , соответствующее значению функции распределения равному p ($F(x_p) = p$). Медиана – квантиль порядка $\frac{1}{2}$.

Определение

Центрированной случайной величиной X , соответствующей величине X называется отклонение случайной величины X ее математического ожидания. (А) Как оценить разброс X относительно среднего? Можно рассмотреть $M(A)$, но он оказывается равен 0. $M(X - m_x) = \sum_{i=1}^n (x_i - m_x) p_i = \sum_{i=1}^n x_i p_i - m_x \sum_{i=1}^n p_i = m_x - m_x = 0$

Поэтому используют $M((X - m_x)^2) = D(X) = Var(X)$

$$D(X) = \sum_{i=1}^n (x_i - m_x)^2 p_i = \sum_{i=1}^n x_i^2 p_i - 2m_x \sum_{i=1}^n x_i p_i + m_x^2 \sum_{i=1}^n p_i = \sum_{i=1}^n x_i^2 p_i - m_x^2$$

$$\text{Для непрерывных величин: } D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - m_x^2$$

Разброс характеризуется средним квадратичным отклонением: $\sigma(x) = \sqrt{D(x)}$

Пример

$$f(x) = \begin{cases} 2x, & x \in [0, 1] \\ 0, & x \notin [0, 1] \end{cases}$$

$$M(X) = \int_0^1 x \cdot 2x dx = \frac{2}{3}$$

$$D(x) = \int_0^1 x^2 \cdot 2x dx - \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

$$\sigma(x) = \frac{1}{3\sqrt{2}} \text{ Чему равно } x_p?$$

$$F(x) = \int_0^\infty 2t dt = x^2, \quad x^2 = p, \text{ следовательно, } x_p = \sqrt{p};$$

$$M_e = x_{\frac{1}{2}} = \frac{1}{\sqrt{2}}$$

$$\text{Начальный момент порядка } S: \alpha_S = \sum_k x_k^s p_k = \int_{-\infty}^{\infty} x^s f(x) dx$$

$$\text{Центральный момент порядка } S: \mu_2 = \sum_k (x_k - M(x))^s p_k = \int_{-\infty}^{\infty} (x - M(x))^s f(x) dx$$

$$\text{Центральный и начальный моменты связаны: } \mu_2 = \alpha_2 - m_x^2$$

$$\begin{aligned} \mu_3 &= \sum_i^n (x_i - m_x)^3 p_i = \sum_i x_i^3 p_i - 3m_x \sum_i x_i^2 p_i + 3m_x^2 \sum_i x_i p_i - m_x^3 \sum_i p_i = \\ &= \alpha_3 - 3m_x \alpha_2 + 2m_x^3 \end{aligned}$$

Можно определить моменты не относительно m_x , а относительно точки a : $M((x-a)^s)$.

Однако центрированные имеют преимущество в части : $D(X) = \min M((x-a)^2)$

$$M(X-a)^2 = \int_{-\infty}^{\infty} (x - m_x + m_x - a)^2 f(x) dx = \int_{-\infty}^{\infty} (x - m_x)^2 f(x) dx + 2(m_x - a) \cdot$$

$$\cdot \int_{-\infty}^{\infty} (x - m_x) f(x) dx + (m_x - a)^2 \int_{-\infty}^{\infty} f(x) dx = D(X) + (m_x - a)^2,$$

следовательно минимум достигается при $m_x = a$

Нечетные центральные моменты характеризуют симметрию распределения (кроме первого, который всегда ноль). Поэтому для характеристики симметрии выбирают третий центральный момент.

$$\text{Коэффициент симметрии: } S_k(x) = \frac{\mu_3}{\sigma^3}.$$

Если симметрия относительно m_x , то $S_k = 0$.

Четвертый момент служит для характеристики крутости, то есть островершинности или плосковершинности распределения. За стандартное распределение, с которым проводится сравнение принято нормальное распределение: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot$

$$e^{-\frac{(x-m_x)^2}{2\sigma^2}}. \text{ Его эксцесс } = E_x = \frac{\mu_4}{\sigma^4} - 3. \text{ Для нормального } E_x = 0.$$

Пример

$$f(x) = \begin{cases} 2x, & x \in [0, 1] \\ 0, & x \notin [0, 1] \end{cases}$$

$$\alpha_3 = 2 \int_0^1 x^4 dx = \frac{2}{5}, \mu_3 = \alpha_3 - 3m_x\alpha_2 + 2m_x^3 = \frac{2}{5} - 3 \cdot \frac{2}{3} \cdot \frac{1}{2} + 2 \cdot \left(\frac{2}{3}\right)^3 = \frac{1}{135}$$

$$S_k = \frac{\mu_3}{\sigma_3} = -\frac{2}{5}\sqrt{2}$$

$$\mu_4 = 2 \int_0^1 x^4 dx = 2 \int_0^1 \left(x - \frac{2}{3}\right)^5 dx + \frac{4}{3} \int_0^1 \left(x - \frac{2}{3}\right)^4 dx = \frac{1}{3^7} - \frac{2^6}{3^7} + \frac{4}{15} \cdot \frac{4^0}{3^5} + \frac{2^7}{5 \cdot 3^6} = \frac{2^2}{3^3} - \frac{2^8}{3^3} + \frac{16}{5 \cdot 3^2} + \frac{2^9}{5 \cdot 3^2} > 0.$$

Помимо важных отдельных моментов существенных значений имеет полный набор моментов (μ_n или α_n). Часто проще найти полный набор, чем само распределение. А при весьма общих условиях набор моментов полностью определяет распределение вероятности.

Теорема

Если две плотности вероятности $f_1(x)$ и $f_2(x)$ – непрерывные случайные величины имеют одинаковые моменты α и функция $f_1(x) - f_2(x)$ представляется сходящимся на $(-\infty; \infty)$ рядом по степеням x , то $f_1(x) = f_2(x)$.

Доказательство:

Пусть $f_1(x) - f_2(x) = c_0 + c_1x + c_2x^2 + \dots =$

$$\int_{-\infty}^{\infty} (f_1(x) - f_2(x))^2 dx = \int_{-\infty}^{\infty} (f_1(x) - f_2(x))(c_0 + c_1x + c_2x^2 + \dots) dx$$

$$= c_0 \int_{-\infty}^{\infty} (f_1(x) - f_2(x)) dx + c_1 \int_{-\infty}^{\infty} (xf_1(x) - xf_2(x)) + c_2 \int_{-\infty}^{\infty} (x^2 f_1(x) - x^2 f_2(x)) + \dots = c_0(1 - 1) + c_1(\alpha_1 - \alpha_1) + c_2(\alpha_2 - \alpha_2) + \dots = 0, \text{ следовательно, } f_1(x) = f_2(x).$$

Замечание

Для "смешанной" случайной величины формулы дискретной и непрерывной объединяются для значений $0 \leq x < 1$ – плотность $f(x) = \frac{1}{2}$.

$$M(X) = 0 \cdot \frac{3}{8} + 1 \cdot \frac{1}{8} + \int_0^1 x \cdot \frac{1}{2} dx = \frac{3}{8}$$

3 Различные законы распределения величин

3.1 Равномерные распределения

Равномерно распределенная непрерывная случайная величина – случайная величина, у которой плотность вероятности постоянна на некотором интервале, а вне него равна нулю.

Пример

Колесо рулетки. Зафиксируем некоторое направление. Тогда угол под которым она останавливается – случайная величина, равномерно распределенная в ин-

тервале $(0, 2\pi)$.

Пример

Производится взвешивание на весах с ценой деления 1 грамм. Если оказывается что вес заключен от k до $k + 1$ грамм, тогда его можно принять за $(k + \frac{1}{2})$ и считать, что при этом допущена ошибка X – случайная величина, равномерно распределенная на интервале $(-\frac{1}{2}, \frac{1}{2})$ грамм.

Пусть распределение X равномерно на (a, b) .

$$f(x) = \begin{cases} c, x \in [a, b] \\ 0, x \notin [a, b] \end{cases}$$

$$\int_a^b c dx = c(b-a) = 1, \text{ значит } c = \frac{1}{b-a}$$

Итого,

$$f(x) = \begin{cases} \frac{1}{b-a}, x \in [a, b] \\ 0, x \notin [a, b] \end{cases}$$

Значит,

$$f(x) = \begin{cases} c, x \in [a, b] \\ 0, x \notin [a, b] \end{cases}$$

Итого,

$$F(x) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, a < x < b \\ 1, x \geq b \end{cases}$$

$$M(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \text{ В силу симметричности } M_e = \frac{a+b}{2} = M(x)$$

Моды это распределение не имеет.

$$D(X) = \frac{1}{b-a} \int_a^b (x - \frac{a+b}{2})^2 dx = \frac{(b-a)^2}{12}$$

$$\sigma(X) = \sqrt{D(X)} = \frac{b-a}{2\sqrt{3}}$$

Распределение симметрично, следовательно асимметрия равна 0 ($S_k = \frac{\mu_3}{\sigma^3} = 0$)

$$\mu_4 = \frac{1}{b-a} \int_a^b (x - \frac{a+b}{2})^4 dx = \frac{(b-a)^4}{80}, \text{ значит } E_x = \frac{\mu_4}{\sigma^4} - 3 = -1.2$$

Вероятность попадания на интервал $(\alpha, \beta) \subset (a, b)$:

$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} \frac{1}{(b-a)} dx = \frac{\beta-\alpha}{b-a}$ – соответствует тому, что используется для вычисления вероятности в методе геометрической вероятности.

3.2 Биномиальные рапределения

Произведение n однотипных опытов, в каждом вероятность события равна p . Случайная величина – число реализаций события.

Случайная величина X принимает целые значения от 0 до n .

$P(X = m) = C_n^m p^m q^n$, $q = (1 - p)$.

1) $\sum_{m=0}^n C_n^m p^m q^{n-m} = 1$ – условие нормирования.

$$M(X) = \sum_{m=0}^n \frac{mn!}{(n-m)!m!} p^m q^{n-m} = np \sum_{m=0}^n \frac{(n-1)!}{(n-m)!(m-1)!} p^{m-1} q^{n-m} = np \sum_{s=0}^{n-1} \frac{(n-1)!}{(n-s-1)!s!} p^s q^{n-s-1} =$$

$$np \sum_{m=0}^n \frac{k!}{(k-s)!s!} p^s q^{k-s} = np$$

$$\alpha_2(X) = \sum_{m=0}^n m^2 P(X = m) = \sum_{m=2}^n m(m-1)P(X = m) + \sum_{m=1}^n mP(X = m) =$$

$$n(n-1)p^2 \sum_{m=2}^n \frac{(n-1)!}{(n-m)!(m-2)!} p^{m-2} q^{n-m} + np = n(n-1)p^2 \sum_{s=0}^{n-2} \frac{(n-2)!}{(n-s-2)!s!} p^s q^{n-s-2} + np =$$

$$n(n-1)p^2 \sum_{s=0}^k \frac{k!}{(k-s)!s!} p^{m-2} q^{k-s} + np = n(n-1)p^2 + np = n(n-1)p^2 + np.$$

$$D(X) = \alpha_2(x) - m_x^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq.$$

$$\sigma(X) = \sqrt{npq}.$$

$$\mu_3(X) = npq(q-p), \text{ следовательно асимметрия } S_k = \frac{npq(q-p)}{(npq)^{\frac{3}{2}}} = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{np(1-p)}}.$$

$$S_k < 0 \text{ если } p > \frac{1}{2}$$

$$S_k = 0 \text{ если } p = \frac{1}{2}$$

$$S_k > 0 \text{ если } p < \frac{1}{2}$$

Если p фиксированно, то $\lim_{n \rightarrow \infty} S_k = 0$ для любого p .

Мода M – целое число, определяемое из двойного неравенства $np - q \leq M \leq np + p$.

Если целое, то две моды: $np + p$ и $np - q$.

3.3 Распределение Пуассона

Пусть нас интересует вероятность того, что за данный промежуток времени произойдет m событий. При этом выполнены следующие условия:

- 1) Произойдет событие или нет в момент времени t не зависит от истории событий, предшествующих моменту t
- 2) Вероятность отдельного события за малый интервал времени Δt возрастает пропорционально длительности интервала, то есть вероятность отдельного события за интервал $(t, t + \Delta t)$ равна $\lambda \Delta t + O(\Delta t)$, $O(\Delta t)$ – бесконечно малое, более высокого порядка малости, чем Δt . λ – среднее число событий на единицу времени (длины)
- 3) Вероятность двух или большего числа событий за $(t, t + \Delta t)$ есть $o(\Delta t)$.

Найдем вероятность того, что в интервале $(0, t)$ не произойдет ни одного события – $P_0(t)$

За промежуток $(0, t + \Delta t)$ не произойдет ни одного события, если не будет событий в интервалах $(0, t)$ и $(t, t + \Delta t)$, то есть, $P_0(t + \Delta t) = P_0(t)(1 - \lambda \Delta t + o(\Delta t))$ следовательно, $\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t) + \frac{o(\Delta t)}{\Delta t}$, следовательно, при $\Delta t \rightarrow 0$ $\frac{dP_0(t)}{dt} = -\lambda P_0(t)$, следовательно $\ln(P_0) = -\lambda t + c$, следовательно $P_0(t) = A e^{-\lambda t}$ При $t = 0$ $P_0(0) = 1$, следовательно, $A = 1$, то есть $P_0(t) = e^{-\lambda t}$

Вероятность того, что за $(0, t)$ произойдет событие $P_1(t)$. Тут две возможности: либо произойдет в $(0, t)$, либо произойдет в $(t, t + \Delta t)$, следовательно $P_1(t + \Delta t) = P_1(t)(1 - \lambda \Delta t + o(\Delta t)) + P_0(t)(\lambda \Delta t + o(\Delta t))$, следовательно $\frac{P_1(t + \Delta t) - P_1(t)}{\Delta t} = -\lambda P_1(t) + \lambda P_0(t) + \frac{o(\Delta t)}{\Delta t}$, следовательно $\frac{dP_1(t)}{dt} = \lambda P_1(t) + \lambda e^{-\lambda t}$.

Его решение: $P_1(t) = \lambda t e^{-\lambda t}$.

Для m событий на $(0, t)$: $\frac{dP_m(t)}{dt} = \lambda P_m(t) + \lambda P_{m-1}(t)$.

Его решение: $P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$.

Это распределение называется распределение Пуассона.

То есть, случайная величина X , распределенная по Пуассону, принимает целые значения от 0 до ∞ с вероятностью $P_m = P(X = m) = \frac{a^m}{m!} e^{-a}$, a – параметр

распределения. $\sum_{m=0}^{\infty} P_m = e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!} = e^{-a} e^a = 1$ (условие нормирования выполнено).

$M(X) = \sum_{m=0}^{\infty} m \frac{a^m}{m!} e^{-a} = \left(\sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} \right) a e^{-a} = a e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} = a e^{-a} e^a = a$.

$\alpha_2(X) = \sum_{m=0}^{\infty} m^2 \frac{a^m}{m!} e^{-a} = a \sum_{m=1}^{\infty} m \frac{a^{m-1}}{(m-1)!} e^{-a} = a \sum_{m=1}^{\infty} ((m-1)+1) \frac{a^{m-1}}{(m-1)!} e^{-a} = a \left(\sum_{m=1}^{\infty} (m-1) \frac{a^{m-1}}{(m-1)!} e^{-a} + \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} e^{-a} \right)$

$= a \left(\sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} e^{-a} + \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} e^{-a} \right) = a(a+1)$

$D(X) = \alpha_2 - m_x^2 = a^2 + a - a^2 = a$. То есть $D(X) = M(X)$.

Теорема

$$\lim_{n \rightarrow \infty} C_n^m p^m (1-p)^{n-m} = \frac{a^m}{m!} e^{-a}$$

Доказательство.

$$C_n^m \frac{a^m}{n^m} (1 - \frac{a}{n})^{n-m} = \frac{(n)(n-1) \dots (n-m+1)}{m!} \frac{a^m}{n^m} \frac{(1 - \frac{a}{n})^n}{(1 - \frac{a}{n})^m} = \frac{(n)(n-1) \dots (n-m+1)}{n^m} \frac{a^m}{m!} \frac{(1 - \frac{a}{n})^n}{(1 - \frac{a}{n})^m}$$

стремится к $\frac{a^m}{m!} e^{-a}$

$$(1 - \frac{a}{n})^n = ((1 - \frac{a}{n})^{\frac{n}{a}})^a \text{ стремится к } e^{-a}$$

Бросание точек на прямую

Пусть

1) Точки распределены статистически равномерно со средней плотностью λ (на единицу длины, площади, объема).

2) точки попадают в неперекрывающиеся области независимым образом.

3) точки появляются по одиночке, а не парами.

Тогда число точек, попавших в область D распределено по Пуассону $P(X = m) = \frac{a^m}{m!} e^{-a}$, где $a = \lambda l (\lambda S_D, \lambda V_D)$

В этом случае говорят, что точки образуют пуассоновское поле.

Вероятность того, что на отрезок l попадет хотя бы одна точка : $P(X \geq 1) = 1 - e^{-a}$

Пример

Число осколков, попадающих в малоразмерную цель при заданном положении точки разрыва, распределение по Пуассону.

Средняя плотность осколков поля равна 3 осколка на квадратный метр. Площадь цели 0.5 метров квадратных. Для поражения цели достаточно попасть в нее хотя-бы одним осколком. Найти вероятность поражения.

$$a = \lambda s = 1.5$$

$$P(X \geq 1) = 1 - e^{-1.5} \approx 1 - 0.223 = 0.777$$

Мода

Легко найти при фиксированном m максимум по a :

$$(a^m e^{-a})' = m a^{m-1} e^{-a} - a^m e^{-a} = 0, \text{ следовательно } m = a.$$

3.4 Показательное распределение

Если число точек попавших на интервал длины t распределено по Пуассону с параметром λt , то расстояние между соседними событиями есть непрерывная случайная величина, распределенная по показательному закону.

$$f(t) = \begin{cases} 0, & t < 0 \\ \lambda e^{-\lambda t}, & t > 0 \end{cases}$$

$$M(X) = \int_0^{\infty} t \lambda e^{-\lambda t} dt = -\lambda t \frac{1}{\lambda} e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}$$

$$D(X) = \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt - \frac{1}{\lambda^2} = -t^2 e^{-\lambda t} \Big|_0^{\infty} - \frac{2}{\lambda} t e^{-\lambda t} \Big|_0^{\infty} - \frac{2}{\lambda^2} e^{-\lambda t} \Big|_0^{\infty} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Пример

Длина свободного пробега (Феллер 2, с. 23)

Пример

Устойчивость неудач (Феллер 2, с. 29).

Показательное распределение связано со временем ожидания (точки – по Пуассону), следовательно, расстояния между ними (время ожидания) – показательное распределение. Рассмотрим очередь. Пусть X_0 – моё время ожидания. Пусть мои друзья подвергли себя опыту того же типа (независимо один от другого). Обозначим их результаты как X_1, X_2, \dots . В каждом случае очереди считаем одинаковыми. (То есть X_0, X_1, \dots – взаимно независимые случайные величины с одним и тем же распределением (например, показательным, хотя это несущественно). Пусть последовательность X_0, X_1, \dots – неограниченная. Чтобы оценить размер моей неудачи, я спрашиваю, как много времени должно пройти, прежде чем один из моих друзей испытает большую неудачу (событие $X_k = X_0$ имеет нулевую вероятность, им пренебрегаем)? То есть вводим время ожидания N как значение первого индекса n , такого, что $X_n > X_0$. Событие $\{N > n - 1\}$ происходит, если и только если максимальный член строки X_0, X_1, \dots, X_{n-1} является начальным – по соображениям симметрии вероятность этого события = $\frac{1}{n}$. Событие $P(N = n) = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}$

Найдем сколько в среднем надо провести испытаний, что бы побить мой рекорд неудач, то есть найдем математическое ожидание N . $M(N) = \sum_{n=1}^{\infty} n \cdot \frac{1}{n(n+1)} = \infty$.

(Объяснение того факта, что всегда кажется, что именно тебе не везет больше всех).

Замечание

Величина X_k не обязательно показательно распределена. Достаточно, что бы они распределены одинаково и были независимы. Тот факт, что это не зависит от закона распределения X_k , используется в статистике для проверки независимости X_k .

3.5 Нормальное распределение

По нормальному закону распределения случайная ошибка.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Выясним смысл параметров σ и μ .

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \text{ Пусть } t = \frac{x-\mu}{\sigma\sqrt{2}}.$$

$$\text{Тогда } M(X) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}t + \mu)e^{-t^2} dt = \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} te^{-t^2} dt + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \cdot 0 +$$

$$\frac{\mu}{\sqrt{\pi}} \cdot \sqrt{\pi} = \mu$$

$$D(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t \cdot 2te^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} (-te^{-t^2}|_{-\infty}^{\infty} +$$

$$\int_{-\infty}^{\infty} e^{-t^2} dt) = \sigma^2$$

$M = Me = \mu$ — мода и медиана совпадают с $M(X)$.

$$\mu_s(x) = \int_{-\infty}^{\infty} (x-\mu)^s f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^s e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \text{ Пусть } t = \frac{x-\mu}{\sigma\sqrt{2}}. \text{ Тогда}$$

$$\mu_s(x) = \frac{(\sigma\sqrt{2})^s}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^s e^{-t^2} dt = \frac{(\sigma\sqrt{2})^s}{\sqrt{\pi}} (-\frac{1}{2}e^{-t^2}t^{s-1}|_{-\infty}^{\infty} + \frac{s-1}{2} \int_{-\infty}^{\infty} t^{s-2} e^{-t^2} dt = \frac{(s-1)(\sigma\sqrt{2})^s}{2\sqrt{\pi}} \int_{-\infty}^{\infty} t^{s-2} e^{-t^2} dt =$$

$$(s-1)\sigma^2\mu_{s-2}, \text{ так как } \mu_{s-2} = \frac{(\sigma\sqrt{2})^{s-2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^{s-2} e^{-t^2} dt$$

То есть имеет место рекуррентное соотношение: $\mu_s = (s-1)\sigma^2\mu_{s-2}$

$\mu_0 = 1$ (по условию нормировки). $\mu_1 = 0$. Пользуясь рекуррентным соотношением, получаем что $\mu_{2k+1} = 0$.

Четные: $\mu_2 = \sigma^2, \mu_4 = 3\sigma^4, \mu_6 = 15\sigma^6 \dots$

Итого: $\mu_{2s} = (2s-1)!!\sigma^{2s}$, то есть все четные центральные моменты однозначно выражаются через σ . Итак, асимптота $S_k = \frac{\mu_3}{\sigma^3} = 0$

$\mu_4 = 3\sigma^4$, следовательно $Ex = \frac{\mu_4}{\sigma^4} - 3 = 0$

Функция нормального распределения

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$\text{Пусть } t = \frac{x-\mu}{\sigma}. \text{ Тогда } F(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Эта функция затабулирована. Иногда вместо нее рассматривают функцию $\phi(x) =$

$$\frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \text{ и } \phi_1(x) = \frac{1}{2\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \Phi^*\left(\frac{\beta-m}{\sigma}\right) - \Phi^*\left(\frac{\alpha-m}{\sigma}\right).$$

Свойства: 1) $\Phi(-\infty) = 0$;

2) $\Phi(\infty) = 1$;

3) $\Phi^*(x)$ – неубывает

4) $\Phi^*(-x) = 1 - \Phi^*(x)$; $\Phi(0) = \frac{1}{2}$

$$\Phi^*(x) = \Phi_1(x) + \frac{1}{2}$$

$$\Phi(x) = 2\Phi_1(\sqrt{2}x)$$

Теорема

$$\text{При } \lim_{x \rightarrow -\infty} 1 - \Phi^*(x) \sim \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

, точнее для всех $x > 0$ справедливо: $(x^{-1} - x^{-3}) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} < 1 - \Phi^*(x) < x^{-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Доказательство:

Очевидно, $(1 - 3x^{-4} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} < \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} < (1 + x^{-2}) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Теперь проинтегрируем это неравенство и получим то что требуется в условии, так как член этого неравенства – произведение членов неравенства из условия теоремы.

"Правило трех сигма:" Для нормального распределения величина

$P(|(X - m)| \leq 3\sigma) = 0,9975 \approx 1$, то есть с практической достоверностью значение нормального распределения случайной величины X попадают в интервал $m - 3\sigma \leq X \leq m + 3\sigma$.

Замечание

X – распределена по нормальному распределению с $m = 10$, $\sigma = 10$. Найти $P(5 < X < 20)$. Какую точность отклонения ϵ случайной величины X от m можно можно гарантировать с вероятностью 0.9.

Решение

$$P(5 < X < 20) = \Phi^*\left(\frac{20-10}{10}\right) - \Phi^*\left(\frac{5-10}{10}\right) = \Phi^*(1) + \Phi^*\left(\frac{1}{2}\right) - 1 = 0.84 - 0.31 = 0.53$$

$$P(|x - 10| < \epsilon) \geq 0.9 \text{ следовательно } P(|(x - 10)| < \epsilon) = 2\Phi^*\left(\frac{\epsilon}{10}\right) - 1 \geq 0.9$$

Из равенства $\Phi^*\left(\frac{\epsilon}{10}\right) = 0.95$, следовательно $\frac{\epsilon}{10} \approx 16.45$, так как функция $\Phi^*(x)$ возрастает, в качестве ϵ можно взять 16.5.

4 Системы случайных величин

Определение

Вектор – функция на пространстве элементарных событий называется случайным вектором.

Замечание

Вектор – функция может иметь значения в пространствах различной размерности.

Координаты случайного вектора в фиксированном базисе – случайные величины. Так что случайный вектор – система случайных величин.

Пример

- 1) Место падения снаряда на плоскости – случайный двумерный вектор.
- 2) Осколок от деления ядра урана характеризуется несколькими параметрами: массой, импульсом, энергией. (это случайный трехмерный вектор).

4.1 Функция распределения системы и плотность вероятности

Для двух случайных величин X, Y .

$F(x, y) = P(X < x, Y < y)$ – это вероятность попадания в угол.

Свойства:

- 1) Формула распределений компонент:

$$F_1(x) = \lim_{y \rightarrow \infty} F(x, y)$$

$$F_2(y) = \lim_{x \rightarrow \infty} F(x, y)$$

- 2) $F(x, y)$ неубывающая функция по каждому аргументу.

$$3) F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0$$

$$4) F(\infty, \infty) = 1$$

Пример

По мишени производится один выстрел, вероятность попадания = $\frac{3}{4}$, X – число попаданий, Y – число промахов, $F(x, y) = ?$.

Дискретная система случайных величин X, Y характеризуется таблицей распределения: $p_{ij} = P(X = x_i, Y = y_j)$

Совместная плотность вероятности системы случайных величин

Для непрерывной случайной величины: $f(x, y) = \lim_{\substack{\Delta y \rightarrow 0 \\ \Delta x \rightarrow 0}} \frac{P((X, Y) \in D)}{\Delta x \Delta y}$ – совместная

плотность вероятности (вероятностное распределение).

Пусть $F(x, y)$ дважды непрерывно дифференцируема.

$$P((X, Y) \in D) = F(x + \Delta x, y + \Delta y) - F(x + \Delta x, y) - F(x, y + \Delta y) + F(x, y)$$

$$f(x, y) = \lim_{\substack{\Delta y \rightarrow 0 \\ \Delta x \rightarrow 0}} \frac{F(x + \Delta x, y + \Delta y) - F(x + \Delta x, y) - F(x, y + \Delta y) + F(x, y)}{\Delta x \Delta y} = \frac{\delta^2 F(x, y)}{\delta x \delta y}$$

Из определения следует, что $P((X, Y) \in \Omega) = \iint_{\Omega} f(x, y) dx dy$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$$

Свойства:

$$1) f(x, y) \geq 0$$

$$2) \iint_{-\infty}^{\infty} f(x, y) dx dy = 1$$

Пример

$$f(x, y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}$$

$$F(x, y) = \frac{1}{\pi^2} \int_{-\infty}^x \int_{-\infty}^y \frac{dx dy}{(1+x^2)(1+y^2)} = \left(\frac{1}{\pi} \arctg x + \frac{1}{2}\right) \left(\frac{1}{\pi} \arctg y + \frac{1}{2}\right)$$

$$P(0 \leq x \leq 1; -1 \leq y \leq 1) = \frac{1}{\pi^2} \int_0^1 dx \int_{-1}^1 dy \frac{dx dy}{(1+x^2)(1+y^2)} = \frac{1}{8}$$

Плотность вероятности компонент
(маргинальные (частные) плотности)

$$F_1(x) = F(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(x, y) dx dy$$

$$f_1(x) = F_1'(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$\text{Аналогично: } f_2(y) = F_2'(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Определим условные плотности вероятности: $f_2(y|x) = \frac{f(x, y)}{f_1(x)}$; $f_1(x|y) = \frac{f(x, y)}{f_2(y)}$

Соотношение $f(x, y) = f_1(x)f_2(y|x)$ соответствует теореме умножения вероятностей. Смысл: $f_2(y|x)$ – это плотность распределения случайной величины Y при условии, что $X = x$.

Для дискретной случайной величины: $p_i = \sum_j p_{ij}$

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

Определение

Случайные величины X, Y называются независимыми, если $f(x, y) = f_1(x)f_2(y)$ (для непрерывных случайных величин)

$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$ (для дискретных случайных величин)

Это означает, что $f_2(y|x) = f_2(y)$ и $f_1(x|y) = f_1(x)$, то есть закон распределения случайной величины Y не зависит от того, какое значение приняла величина X .

Пример

$f(x, y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}$, $f_1(x) = \frac{1}{\pi(1+x^2)}$, $f_2(y) = \frac{1}{\pi(1+y^2)}$ – независимые случайные величины.

Функции распределения и плотности систем из n случайных величин

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

Свойства те же, что и при $n = 2$.

$$f(x_1, x_2, \dots, x_n) = \frac{\delta^n F(x_1, x_2, \dots, x_n)}{\delta x_1 \delta x_2 \dots \delta x_n}$$

$$F_1(x_1) = F(x_1, \infty, \dots, \infty)$$

$$f_1(x_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \dots dx_n$$

Случайные величины независимы, если $f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \dots f_n(x_n)$

4.2 Числовые характеристики системы из 2х случайных величин

Основные характеристики системы (X, Y) :

$M(X), M(Y), D(X), D(Y), K_{XY}$ – они уже определены, но их можно находить, не вычисляя $f_1(x), f_2(y)$.

Дискретная случайная величина	Непрерывная случайная величина
$M(X) = \sum_{i,j} x_i p_{i,j}$	$\iint_{-\infty}^{\infty} x f(x, y) dx dy$
$M(Y) = \sum_{i,j} y_j p_{i,j}$	$\iint_{-\infty}^{\infty} y f(x, y) dx dy$
$D(X) = \sum_{i,j} (x_i - M(X))^2 p_{i,j}$	$\iint_{-\infty}^{\infty} (x - M(X))^2 f(x, y) dx dy$ $= \iint_{-\infty}^{\infty} x^2 f(x, y) dx dy$
$K_{X,Y} = \sum_{i,j} (x_i - M(X))(y_j - M(Y)) p_{i,j}$ $= cov(X, Y)$	$\iint_{-\infty}^{\infty} (x - M(X))(y - M(Y)) f(x, y) dx dy$

Теорема

Корреляционный момент независимых случайных величин равен нулю

Доказательство:

Для непрерывных случайных величин:

Они независимы, значит, $f(x, y) = f_1(x)f_2(y)$

$$K_{XY} = \int_{-\infty}^{\infty} (x - M(X)) f_1(x) dx \int_{-\infty}^{\infty} (y - M(Y)) f_2(y) dy = 0$$

$$r_{XY} = \frac{K_{XY}}{\sigma_x \sigma_y}, \sigma_x = \sqrt{D(X)}, \sigma_y = \sqrt{D(Y)}.$$

Пример

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 < 1 \\ 0, & x^2 + y^2 > 1 \end{cases}$$

Величины являются зависимыми (если $X = 0$, то Y может с равными вероятностями принимать значения на $[-1, 1]$, если $x = 1$, то Y может принять одно единственное значение $Y = 0$.)

$M(X) = M(Y) = 0$ по соображениям симметрии.

$$K_{XY} = \iint_C xf(x, y)dxdy = \frac{1}{\pi} \iint_C xydxdy = \frac{1}{\pi} \iint_{C_1} xydxdy + \frac{1}{\pi} \iint_{C_2} xydxdy + \frac{1}{\pi} \iint_{C_2} xydxdy + \frac{1}{\pi} \iint_{C_3} xydxdy + \frac{1}{\pi} \iint_{C_4} xydxdy = 0.$$

То есть независимые следовательно некоррелирующие, обратное неверно.

$$-1 \leq r_{xy} \leq 1$$

Если $Y = aX + b$, то $r_{XY} \pm 1$

Если $a > 0$, то $+1$, если $a < 0$ то -1

Докажем это позже.

Пример

а) Вес и рост человека связаны положительной корреляцией

б) Возраст взрослого человека и количество волос у него на голове связаны отрицательной корреляцией

$$\alpha_{k,s}(X, Y) = \sum_{i,j} x_i^k y_j^s p_{ij} = \iint_{-\infty}^{\infty} x^k y^s f(x, y)dxdy - \text{начальный момент корреляции.}$$

$$\mu_{k,s}(X, Y) = \sum_{i,j} (x_i - M(X))^k (y_j - M(Y))^s p_{ij} = \iint_{-\infty}^{\infty} (x - M(X))^k (y - M(Y))^s f(x, y)dxdy$$

Числовые характеристики случайных величин

X_1, X_2, \dots, X_n - их характеристики: n математических ожиданий (m_1, m_2, \dots, m_n) ,

n дисперсий (D_1, \dots, D_n) , $n(n-1)$ корреляционных моментов:

$$K_{ij} = M((X_i - m_i)(X_j - m_j)) \quad (D_i = K_{ii})$$

Корреляционные матрицы K

$$\begin{pmatrix} D_1 & K_{12} & K_{13} & \dots & K_{1n} \\ K_{21} & D_2 & K_{23} & \dots & K_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \vdots & \vdots & D_n \end{pmatrix}$$

Если X_1, X_2, \dots, X_n - не коррелированы, то корреляционная матрица

$$\begin{pmatrix} D_1 & 0 & 0 & \dots & 0 \\ 0 & D_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \vdots & D_n \end{pmatrix}$$

$$\text{Коэффициент корреляции } r_{ij} = \frac{k_{ij}}{\sigma_i \sigma_j}$$

Нормированная корреляционная матрица $\begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{n1} & r_{n2} & \cdot & \cdot & 1 \end{pmatrix}$ – характеризует именно коррелированность, безотносительно к их рассеиванию.

4.3 Нормальный закон распределения систем случайных величин

(X, Y) – нормированные случайные величины.

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)} \left(\frac{(x-m_x)^2}{\sigma_x^2} + \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2} \right)}$$

Можно показать, что m_x, m_y – математические ожидания, $\sigma_x\sigma_y$ – среднеквадратичные отклонения, r – коэффициент корреляции X и Y .

Теорема

Для нормального распределения случайной величины X и Y независимость эквивалентна некоррелированности.

Доказательство:

Из независимости следует некоррелируемость для любой случайной величины.

Пусть X, Y – нормированны, и $r = 0$, следовательно $f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}} = f_1(x)f_2(x)$, то есть X и Y – независимы.

Регрессия

Рассмотрим условную плотность вероятности $f_2(y|x)$. Введем условное математическое ожидание: $M(Y|X = x) = \int_{-\infty}^{\infty} y f_2(y|x) dy = \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y f(x, y) dy$

$$M(Y|X = x) = \int_{-\infty}^{\infty} y f_2(y|x) dy = \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y f(x, y) dy$$

,в предположении, что интеграл сходится абсолютно.

$M(Y|X)$ – это случайная величина (она является функцией только от X , а не от Y , которая называется регрессией Y на X).

Это соотношение не определено, когда $f_1(x) = 0$. Пусть в этом случае $M(Y|X = x) = 0$, тогда $M(Y|X = x)$ определено, когда $f_1(x) = 0$. Пусть в этом случае $M(Y|X = x) = 0$, тогда $M(Y|X)$ определено для любой непрерывной плотности распределения. Можно ввести аналогичным образом условную дисперсию $Var(Y|X) = D(Y|X)$.

$$\text{Для нормального закона распределения: } f_2(y|x) = \frac{1}{\sigma_y \sqrt{1-r^2} \sqrt{2\pi}} e^{-\frac{1}{2(1-r^2)} \cdot (\frac{(x-m_x)^2}{\sigma_x^2} - r \frac{(y-m_y)^2}{\sigma_y^2})} =$$

$$\frac{e^{-\frac{1}{2(1-r^2)} \cdot (\frac{(x-m_x)^2}{\sigma_x^2} - r \frac{(y-m_y)^2}{\sigma_y^2})}}{\sigma_y \sqrt{1-r^2} \sqrt{2\pi}}$$

То есть это плотность нормального закона с $m_{y|x} = m_y + r \frac{\sigma_y}{\sigma_x} (x - m_x)$ и средне-квадратичное отклонения $\sigma_{y|x} = \sigma_y \sqrt{1-r^2}$

То есть от x в условном законе зависит лишь $m_{y|x}$, но не $\sigma_{y|x}$. Откладывая $m_{y|x}$ на оси ординат при различных x получим линию регрессии $y = m_y + r \frac{\sigma_y}{\sigma_x} (x - m_x)$

Пример

Гальтона – характеризует эмпирическое значение формул линейной регрессии. Пусть X_1 и X_2 – характеризуют рост (измеренный по отклонению от среднего в сантиметрах) соответственно отцов и сыновей в некоторой популяции. Рост случайно выбранного сына – нормальная случайная величина с $M(X_2) = 0$ и $D(X_2) = \sigma_2^2$, отца – с $M(X_1) = 0$, $D(X_1) = \sigma_1^2$. Но в подпопуляции сыновей, отцы которых имеют фиксированный рост x , рост сыновей представляется нормальной величиной с (по формулам регрессии) математическим ожиданием $r \frac{\sigma_2}{\sigma_1} x$ и дисперсией $\sigma_2^2(1-r^2) < \sigma_2^2$.

Таким образом, регрессия X_2 на X_1 показывает как много статистической информации относительно X_2 содержится в наблюдении над X_1 .

5 Функции случайных величин

5.1 Числовые характеристики функций случайных величин

$$Y = \phi(X), M(Y) = M(\phi(X))$$

$$\text{Для непрерывной случайной величины: } M(\phi(X, Y)) = \int_{-\infty}^{\infty} \phi(x) f(x) dx$$

$$\text{Для дискретной случайной величины: } M(\phi(X, Y)) = \sum_{i,j} \phi(x_i, y_j) p_{i,j}$$

Пример

На плоскости задан отрезок длины l , вращающаяся случайным образом так, что все направления одинаково вероятны. Найти среднее значение длины проекции отрезка на неподвижную ось.

$$Y = l|\cos \alpha|, \alpha \text{ распределенна равномерно от } 0 \text{ до } 2\pi$$

$$M(l|\cos\alpha|) = \int_0^{2\pi} l|\cos\alpha| \frac{d\alpha}{2\pi} = \frac{2l}{\pi} \int_0^{2\pi} \cos\alpha d\alpha = \frac{2l}{\pi}$$

$$D(l|\cos\alpha|) = \frac{4}{2\pi} \int_0^{\frac{\pi}{2}} l^2 \cos^2\alpha d\alpha - \frac{4l^2}{\pi^2} = \frac{2l^2}{\pi} \int_0^{\frac{\pi}{2}} (\frac{1}{2} + \frac{1}{2}\cos 2\alpha) d\alpha - \frac{4l^2}{\pi^2} = \frac{l^2}{2} - \frac{4l^2}{\pi^2}$$

Пример

В процессе слежения радиолокатором за определенным объектом пятно изображающее объект, все время удерживается в пределах экрана (круг радиуса R) и занимает на нем случайное положение с постоянной плотностью вероятности. Найти среднее расстояние до центра экрана.

$$f(x, y) = \frac{1}{\pi R^2}$$

$$M(r) = \iint_K \sqrt{x^2 + y^2} \frac{dx dy}{\pi R^2} = \frac{1}{\pi R^2} \int_0^{2\pi} d\phi \int_0^R \rho^2 d\rho = \frac{l}{3} R$$

Теоремы о числовых характеристиках

$$1) M(C) = C$$

$$2) D(C) = 0$$

$$3) M(cX) = cM(X)$$

Доказательство:

$$M(cX) = \int_{-\infty}^{\infty} cxf(x)dx = c \int_{-\infty}^{\infty} xf(x)dx = xM(X)$$

$$4) D(cX) = c^2 D(X)$$

Доказательство:

$$\overline{D(cX)} = \overline{M((cX - M(cX)))} = \overline{M((cX - cM(X))^2)} = \overline{M(c^2(X - M(X))^2)} = c^2 \overline{M((X - M(X))^2)} = c^2 D(X)$$

$$5) M(X + Y) = M(X) + M(Y), M(\sum x_i) = \sum M(x_i)$$

Доказательство:

$$\begin{aligned} M(X+Y) &= \iiint_{-\infty}^{\infty} (x+y)f(x,y)dx dy = \iiint_{-\infty}^{\infty} xf(x,y)dx dy + \iiint_{-\infty}^{\infty} yf(x,y)dx dy = \int_{-\infty}^{\infty} dx x \int_{-\infty}^{\infty} f(x,y)dy \\ &+ \int_{-\infty}^{\infty} dy y \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{\infty} dx x f_1(x) + \int_{-\infty}^{\infty} dy y f_1(y) = M(X) + M(Y) \end{aligned}$$

$$6) M(\sum a_i X_i + b) = \sum a_i M(x_i) + b$$

$$7) D(X + Y) = D(X) + D(Y) + 2K_{XY}$$

Доказательство:

$Z = X + Y$. Перейдем к центрированным величинам.

$$8) D\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n (a_i^2 D(X_i) + 2 \sum_{i < j}^n a_i a_j K_{i,j})$$

Доказательство:

$$Y_i = a_i X_i,$$

$$\sum_{i=1}^n a_i X_i + b = \sum_{i=1}^n Y_i + b, \text{ следовательно } D\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n D(Y_i) + \sum_{i < j} K_{i,j}^y =$$

$$\sum_{i=1}^n a_i^2 D(X_i) + 2 \sum_{i < j} K_{i,j}^y$$

5.2 Законы распределения функций случайной величины

$Y = \phi(x)$. Пусть ϕ непрерывна и дифференцируема. X – непрерывная случайная величина. Пусть значения случайной величины X лежат на (a, b) , и ϕ на (a, b) монотонна.

а) ϕ монотонно возрастает на (a, b) . $f(x)$ – плотность вероятности X , $g(y)$ – плотность вероятности Y , $G(y)$ – функция распределения Y . $G(y) = P(Y < y) = P(a < X < x)$ (так как ϕ – возрастающая функция.)

$$P(a < X < x) = \int_a^x f(x) dx.$$

Пусть $\psi(y)$ – обратная функция к $\phi(x)$. Тогда $G(Y) = \int_a^{\psi(y)} f(x) dx$. $g(y) = G'(y) =$

$$f(\psi(y))\psi'(y)$$

б) ϕ – монотонно убывает на (a, b)

$$G(y) = P(Y < y) = P(x < X < b) = \int_{\psi(y)}^b f(x) dx$$

$$g(y) = G'(y) = -f(\psi(y)).$$

Общая формула: $g(y) = f(\psi(y))|\psi(y)|$

Пример

Пушка, расположенная на расстоянии l от бесконечной прямолинейной стены, равномерно вращается. В произвольный момент времени при каждом оборо-

те производится выстрел. Найти плотности распределения точек попаданий на стене.

$x = l \cdot tg\phi$. Функция распределения равномерна на $(-\frac{\pi}{2}, \frac{\pi}{2})$:

$$f(\phi) = \begin{cases} \frac{1}{\pi}, \phi \in [-\frac{1}{\pi}; \frac{1}{\pi}] \\ 0, \phi \notin [-\frac{1}{\pi}; \frac{1}{\pi}] \end{cases}$$

$g(x) = \frac{1}{\pi} |(arctg \frac{x}{l})'| = \frac{l}{\pi(l^2+x^2)}$ – распределение Коши.

Пусть ϕ немонотонно дифференцируемая функция на (a, b) .

$$G(y) = P(Y < y) = P(X \in \Delta_1(y)) + P(X \in \Delta_2(y)) \dots = \sum_i \int_{\Delta_i(y)} f(x) dx$$

Пример

Для получения формулы для $g(y)$ оказывается удобнее действовать по-другому. Разобьем (a, b) на интервалы монотонности ϕ .

$$G(y) = P(Y < y) = P((Y < y) \cap (x \in (a, a_1))) + P((Y < y) \cap (x \in (a_1, a_2))) + \dots + P((Y < y) \cap (x \in (a_{n-1}, b)))$$

$g(y) = G'(y)$. Для каждого слагаемого производная находится по формуле для монотонной функции. Пусть на i_m интервале монотонности обратная функция

$$\text{к } \phi(x) - \text{это } \psi_i(y). \text{ Тогда } g(y) = \sum_{i=1}^n f(t_i(y)) |\psi(y)|$$

Пример

$$f(x) = \frac{1}{\pi(1+x^2)}, Y = X^2, X = \pm\sqrt{Y} \text{ (два интервала монотонности)}$$

$$\begin{cases} g(y) = \frac{1}{\pi(1+y)} \left| \frac{1}{2\sqrt{y}} \right| + \frac{1}{\pi(1+y)} \left| -\frac{1}{2\sqrt{y}} \right| = \frac{1}{\pi(1+y)} \left| \frac{1}{\sqrt{y}} \right|, y > 0 \\ g(y) = 0, y \leq 0 \end{cases}$$

Для дискретной случайной величины $Y = \phi(x)$ надо найти все $y_i = \phi(x_i)$ и если среди y_i есть совпадающие, то надо сложить соответствующие вероятности.

6 Предельные теоремы теории вероятностей

Предельные теоремы связаны с большим числом однородных опытов или с большим числом складывающихся случайных воздействий. Суть "закона больших чисел" в том, что при большом числе случайных явлений средний их результат стабилизируется и практически перестает быть случайным.

Неравенство Чебышева

Пусть X – случайная величина с математическим ожиданием m_x , дисперсией D_x .
Тогда $P(|X - m_x| \geq \alpha) \leq \frac{D_x}{\alpha^2}$

Доказательство:

Для непрерывных случайных величин.

$$\begin{aligned} D_x &= \int_{-\infty}^{\infty} (x - m_x)^2 f(x) dx \geq \int_{|x-m_x|>\alpha} |x - m_x|^2 f(x) dx \geq \\ &\geq \alpha^2 \int_{|x-m_x|>\alpha} f(x) dx = \alpha^2 P(|X - m_x| \geq \alpha) \end{aligned}$$

Для дискретных случайных величин – аналогично.

Пример

$$P(|X - m_x| \geq 3\sigma) \leq \frac{D_x}{9\sigma_x^2} = \frac{1}{9}$$

Видно, что оценка по приближению Чебышева завышена. Например для нормального распределения это $0.003 < < \frac{1}{9}$

Пусть X – случайная величина с m_x и D_x

X_1 – значение X в первом опыте. X_2 – во втором и так далее. $Y = \frac{\sum_{i=1}^n x_i}{n}$

$$M(Y) = \frac{1}{n} \sum_{i=1}^n M(X_i) = \frac{1}{n} n m_x = m_x$$

$$D(Y) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{D_x}{n}$$

Теорема (Чебышева)

При неограниченном увеличении числа независимых опытов среднее арифметическое наблюдаемых значений случайной величины, имеющих ограниченную дисперсию D_x , сходится по вероятности к её математическому ожиданию, то

есть $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n x_i - m_x\right| < \varepsilon\right) = 1$, где ε – сколь угодно маленькое положительное число.

Доказательство:

Y имеет $m_y = m_x$, $D_y = \frac{1}{n} D_x$ Неравенство Чебышева для Y : $P(|Y - m_y| \geq \varepsilon) \leq \frac{D_y}{\varepsilon^2} = \frac{D_x}{n\varepsilon^2}$. Пусть ε и Δ сколь угодно малые. Выберем n так, что $\frac{D_x}{n\varepsilon^2} < \Delta$.

Тогда $P\left(\left|\frac{1}{n} \sum_{i=1}^n x_i - m_x\right| \geq \varepsilon\right) < \Delta$, переходим к противоположному событию,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n x_i - m_x\right| < \varepsilon\right) > 1 - \Delta.$$

Теорема (Бернулли)

При неограниченном увеличении числа опытов n частота события A сходится по вероятности к его вероятности p .

Доказательство:

Частота $P^* = \frac{m}{n}$. X_1 – число появлений события A в одном опыте, X_2 – в двух и так далее. Закон распределения для всех X_i – одинаков:

x_i	0	1
p_i	q	p

$$M(X_i) = P, D(X_i) = pq.$$

$$P^* = \frac{1}{n}\sum_i^n x_i, \text{ следовательно, по теореме Чебышева } P(|P^* - p| < \varepsilon) > 1 - \delta$$

Замечание

Математическое ожидание и дисперсия биномиального распределения находятся проще с помощью системы случайных величин $X = \sum_{i=1}^n X_i$, $M(X_i) = p$,

$$D(X_i) = pq.$$

$$M\left(\sum_{i=1}^n X_i\right) = np, D\left(\sum_{i=1}^n X_i\right) = npq - \text{ для независимых случайных величин}$$

Центральная предельная теорема

Если X_1, X_2, \dots, X_n – независимые случайные величины, имеющие один и тот же закон распределения с математическим ожиданием m и дисперсией σ^2 , то при неограниченном увеличении n закон распределения $Y_n = \sum_{i=1}^n X_k$ неограниченно приближается к нормальному.

Частный случай этой теоремы – теорема Муавра-Лапласа.

Если произведение n независимых опытов, в каждом из которых событие A появляется с вероятностью p , то справедливо соотношение: $P(\alpha < Y < \beta) \approx \Phi^*\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi^*\left(\frac{\alpha - np}{\sqrt{npq}}\right)$

Y – число появлений события A в опытах $q = 1 - p$.

Пример

60% изготовленных изделий – первого сорта. Приемщик берет наугад 200 изделий. Чему равна вероятность того, что среди них от 120 до 150 изделий первого сорта?

$$P(120 < Y < 150) \approx \Phi^*\left(\frac{150-120}{\sqrt{48}}\right) - \Phi^*\left(\frac{120-120}{\sqrt{48}}\right) = \Phi^*(4.33) - \Phi^*(0) \approx 0.5$$

7 Математическая статистика

7.1 Гистограмма. Статистическая функция распределения.

Математическая статистика имеет дело с анализом экспериментальных данных, получаемых в результате наблюдения массовых случайных явлений.

Основные задачи:

- 1) Задача определения закона распределения случайной величины по статистическим данным. В связи с ограниченностью экспериментального материала необходимо научиться отбирать в нем существенные и отбрасывать то, что не присуще закону распределения.
- 2) Задача проверки правдоподобия гипотез. Например, согласуются ли экспериментальные данные с тем, что случайная величина распределена по данному закону, или существует или нет зависимость между двумя случайными величинами.
- 3) Задача отыскания неизвестных параметров распределения. Часто закон распределения находить не требуется, или он известен заранее. Требуется находить лишь числовые характеристики.

Пусть надо определить закон распределения некоторой случайной величины X . Производится ряд измерений величины X . Совокупность полученных результатов называется простой статистической совокупностью (выборкой). Происхождение названия: часто в статистике рассматривается распределение какого-то признака среди большой группы объектов (Например вес коровы).

Множество всех объектов при этом называется генеральной совокупностью. Для оценки признака выбирают часть этой совокупности (её называют выборкой). Например, Остап Бендер изучал распределение бриллиантов в стульях и проводил выборку из генеральной совокупности.

Определение

Статистическая функция распределения случайной величины X – это частота события $X < x$ в данной выборке: $F^*(x) = P^*(X < x)$. Это ступенчатая функция. По теореме Бернулли, при n стремящемся к бесконечности, частота событий стремится к вероятности, $F^*(x)$ сходится по вероятности к $F(x)$. При большом объеме (n) выборки строить так $F^*(x)$ трудоёмко. Там действуют так: весь интервал полученных значений X разбивают на разряды (интервалы). Подсчитывают число наблюдений, приходящихся на данных интервал m_i и находят

частоту $p_i^* = \frac{m_i}{n}$.

I_i	(x_1, x_2)	(x_2, x_3)	\dots	(x_k, x_{k+1})
p_i^*	p_1^*	p_2^*	\dots	p_k^*

Эта таблица называется статистическим рядом. Длины разрядов можно брать и одинаковыми, и различными. Графически, статистический ряд оформляется в виде гистограммы. На каждом интервале откладывается вверх прямоугольник высоты $\frac{p_i^*}{l_i}$, где l_i – длина интервала. (на длину интервала можно не делить , если они все одинаковые). Площадь гистограммы равна единице. Если число опытов и число разрядов увеличивать, то гистограмма буде стремиться к плавной кривой – графику плотности вероятности случайной величины X .

Для построения статистической функции распределения обычно приписывают p_i^* середине i -го интервала $\frac{x_i+x_{i+1}}{2} = x_i^*$. И дальше строят ступенчатую функцию для данного статистического ряда. Иногда строят не ступенчатую, а например функцию соединяя значения, соответствующие концам интервала

Можно ввести по аналогии с вероятностными – статистические числовые характеристики. Однако, в большей степени они интересны не сами по себе, а той информацией, которую они содержат об измеряемой величине.

7.2 Точечные оценки параметров распределения

Часто вид распределения известен заранее, а определить надо его параметры (например для нормального m и σ), или закон распределения нас не интересует, а нужны лишь его числовые характеристики. Если при этом после статистической обработки мы предлагаем для данного параметра в качестве приближения одно число, то это число называется точечной оценкой параметра. Пусть у случайной величины X закон распределения содержит параметр a . Мы производим n независимых измерений величины X и получаем значения X_1, X_2, \dots, X_n . Их можно рассматривать как n экземпляров случайной величины X , то есть n независимых случайных величин с тем же законом распределения. На основе X_1, X_2, \dots, X_n мы находим оценку \hat{a} параметра $a = \hat{a} = \hat{a}(X_1, X_2, \dots, X_n)$, то есть оценка это случайная величина. Ее закон распределения зависит от закона распределения X и от числа опытов. Какие оценки хороши?

Определение

Оценка \hat{a} называется состоятельной, если при $n \rightarrow \infty$ она сходится по вероятности к a

Определение

Оценка \hat{a} называется несмещенной, если ее математическое ожидание равно a . $M(\hat{a}) = a$. Это говорит об отсутствии систематических ошибок при оценивании.

Определение

Несмещенная оценка называется эффективной, если её дисперсия минимальна по сравнению с другими оценками.

Оценки математического ожидания и дисперсии

Пусть случайная величина X имеет $M(X) = m$ и $D(X) = D$ (M и D неизвестны). X_1, X_2, \dots, X_n – наблюдаемые величины X .

$$\hat{m} = \frac{\sum_{i=1}^n x_i}{n}$$

Это состоятельная оценка так как по закону больших чисел, при $n \rightarrow \infty$ \hat{m} схо-

дится по вероятности к m . Оценка \hat{m} несмещенная, так как $M(\hat{m}) = \frac{\sum_{i=1}^n M(X_i)}{n} =$

m . Дисперсия оценки: $D(\hat{m}) = \frac{1}{n^2} \sum_{i=1}^n D = \frac{D}{n}$

Отсюда видно, зачем надо производить несколько измерений, а потом усреднять – уменьшать дисперсию (точность увеличивается в \sqrt{n} раз).

Эффект оценки зависит от вида распределения. Если X распределена нормально, то оценка \hat{m} эффективна. Для других законов это может не выполняться. Если n велико (практически $n > 30$) и x_i некоррелируют, то по центральной предельной теореме \hat{m} подчиняется нормальному распределению со средним значением m и среднеквадратичным отклонением $\sqrt{\frac{D}{n}}$

Оценка дисперсии.

$$\text{Можно предположить: } S^2 = \frac{\sum_{i=1}^n (X_i - \hat{m})^2}{n}, \quad \hat{m} = \frac{\sum_{i=1}^n x_i}{n}$$

1) Оценка состоятельна

$$S^2 = \frac{\sum_{i=1}^n (X_i)^2}{n} - \hat{m}^2$$

Это среднее арифметическое n наблюдений значений величины X^2 . Он сходится по вероятности (по закону больших чисел) к $M(X^2) = \alpha_2(X)$. Второй член сходится по вероятности к m^2 . То есть $S^2 \xrightarrow{n \rightarrow \infty} \alpha_2(x) - m^2 = D$ (по вероятности)

2) Оценка S^2 – смещена.

$$S^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \frac{\sum_{i=1}^n x_i^2}{n^2} - 2 \frac{\sum_{i < j} x_i x_j}{n^2} = \frac{n-1}{n} \sum_{i=1}^n x_i^2 - 2 \frac{\sum_{i < j} x_i x_j}{n^2}$$

$$M(S^2) = \frac{n-1}{n^2} \sum_{i=1}^n M(X_i^2) - \frac{2}{n^2} \sum_{i < j} M(X_i X_j)$$

S^2 не зависит от выбора начала координат. Пусть начало координат в точке m .

Тогда $M(X_i^2) = M(\hat{X}_i^2) = D$, $\sum_{i=1}^n M(X_i^2) = nD$, $M(X_i X_j) = M(\hat{X}_i \hat{X}_j) = K_{ij} = 0$.

(так как опыты независимы, следовательно некоррелируют).

Значит, $M(S^2) = \frac{n-1}{n} D \neq D$

Что бы оценка стала несмещенной, ее надо умножить на $\frac{n}{n-1}$.

$$\hat{D} = \frac{\sum_{i=1}^n ((x_i - \hat{m})^2)}{n-1} - \text{несмещенная оценка. (или } \hat{D} = \frac{\sum_{i=1}^n x_i^2}{n-1} - \hat{m}^2 \frac{n}{n-1} \text{)}$$

При больших значениях n — S^2 и \hat{D} различаются мало. При известном математическом ожидании (m) несмещенная оценка дисперсии: $\hat{D} = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$.

(То есть $\frac{1}{n-1}$ происходит из-за того, что была оценкой \hat{m} , а не m).

$$M\left(\frac{1}{n} \sum_{i=1}^n (X_i - m)^2\right) = \frac{1}{n} \sum_{i=1}^n M((X_i - m)^2) = \frac{1}{n} \sum_{i=1}^n D = D, \text{ то есть она несмещенная.}$$

Пример

Для определения точности измерительного прибора (лазерного дальномера) систематическая ошибка которого практически равна нулю, было произведено 5 независимых измерений.

$x_i = 2781$ метра, 2836 метра, 2807 метра, 2763 метра, 2858 метра. Определить несмещенную оценку дисперсий ошибок измерительного прибора, если значение измеряемой величины:

а) известно и равно 2800 м

б) неизвестно

Решение

$$\text{а) } \hat{D}(X) = \frac{1}{n} \sum_{j=1}^n (x_j - m)^2 = \frac{6439}{5} = 1287.8 \text{ квадратных метров}$$

$$\text{б) } \hat{m} = \frac{1}{n} \sum_{j=1}^n (x_j) = 2809 \text{ метров}$$

$$\hat{D}(X) = \frac{1}{n} \sum_{j=1}^n (x_j - \hat{m})^2 = \frac{6034}{4} = 1508.5 \text{ квадратных метров}$$

7.3 Интервальные оценки параметров

Если для оценки параметра предлагается интервал, в который его значение попадает с заданой вероятностью, то такая оценка называется интервальной.

Определение

Доверительным интервалом называется интервал, который с заданой доверительной вероятностью β накрывает оцениваемый параметр a .

Оценка \hat{a} случайна, и поэтому замена a на \hat{a} может привести к ошибкам. Интервалы оценки вводятся для того, что бы можно было судить о надежности оценки. Для симметричного доверительного интервала его ширина 2ε определяется условием: $P(|\hat{a} - a| \leq \varepsilon) = \beta$

Доверительный интервал: $(\hat{a} - \varepsilon, \hat{a} + \varepsilon) = I_\beta$

Замечание

Здесь a не является случайным, случаен интервал, положение которого задается случайным положением \hat{a} . $\hat{a} - \varepsilon$ и $\hat{a} + \varepsilon$ – доверительные границы.

Пусть проведено n измерений величины X (с $M(X) = m$, $D(X) = D$)

$$\text{Оценки: } \hat{m} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{D} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m})^2$$

Построим доверительный интервал I_β для математического ожидания \hat{m} – сумма n независимых одинаково распределенных случайных величин X_i .

По центральной предельной теореме при большом n закон распределения \hat{m} близок к нормальному с математическим ожиданием m и дисперсией $\frac{D}{n}$. Пусть дисперсия D известна. Найдем доверительные границы $P(|\hat{m} - m| < \varepsilon_\beta) = \beta$.

Так как \hat{m} распределен нормально с $\sigma_{\hat{m}} = \sqrt{\frac{D}{n}}$

$$\beta = P(|\hat{m} - m| < \varepsilon_\beta) = 2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\hat{m}}}\right) - 1.$$

$$\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Находим t_β по таблице функции $\Phi^*(x)$: $2\Phi^*(t_\beta) - 1 = \beta$

$$\Phi^*(t_\beta) = \frac{\beta+1}{2}$$

$$\text{Тогда, } \varepsilon_\beta = t_\beta \sqrt{\frac{D}{n}}$$

$$\text{Доверительный интервал: } (\hat{m} - t_\beta \sqrt{\frac{D}{n}}, \hat{m} + t_\beta \sqrt{\frac{D}{n}})$$

Если D неизвестна, то в качестве её ориентировочного значения можно взять \hat{D} и положить приближенно $\sigma_{\hat{m}} = \sqrt{\frac{\hat{D}}{n}}$

Но при этой замене возникает дополнительная ошибка. Для точного определения доверительного интервала необходимо знать вид закона распределения X . Пусть распределение распределено по нормальному закону. Произведено n измерений. $\hat{m} = \frac{1}{n} \sum_{i=1}^n X_i$, $\hat{D} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m})^2$.

Оказывается, что величина $T = \sqrt{n} \frac{\hat{m} - m}{\sqrt{\hat{D}}}$ подчиняется закону распределения Стьюдента с $(n - 1)$ степенями свободы, не зависящих от неизвестных параметров m и D , плотность которого $S_{n-1} = \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}} (1 + \frac{t^2}{n-1})^{-\frac{n}{2}}$, где $\Gamma(x)$ – гамма

– функция Эйлера. ($\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$.)

Найдем доверительный интервал соответствующий вероятности β :

$$P(|\hat{m} - m| < \varepsilon) = \beta \Leftrightarrow P\left(\frac{|\sqrt{n}\hat{m} - m|}{\sqrt{\hat{D}}} < \frac{\varepsilon_\beta}{\sqrt{\frac{\hat{D}}{n}}}\right) = \beta \Leftrightarrow P(|T| < \frac{\varepsilon_\beta}{\sqrt{\frac{\hat{D}}{n}}}) = \beta.$$

Найдем t_β из : $P(|T| < t_\beta) = \beta$

$$P(|T| < t_\beta) = \int_{-t_\beta}^{t_\beta} S_{n-1}(t) dt = \beta.$$

(Замечание: S_{n-1} – четная функция, поэтому $\int_{-t_\beta}^{t_\beta} S_{n-1}(t) dt = 2 \int_0^{t_\beta} S_{n-1}(t) dt$)

$t_\beta(n)$ называется коэффициент Стьюдента. Для них существуют таблицы.)

Пример

Оцениваем средний рост мужчины. Выборка 10 человек. Результат: 160, 160, 167, 170, 173, 176, 178, 178, 181, 181.

Найдем $\hat{m} = 172.4$. $\hat{D} = \frac{1}{9} \sum_{i=1}^{10} (x_i - 172.4)^2 = 62.9$

$\sqrt{\frac{\hat{D}}{n}} = 2.51$. Если $\beta = 0.95$, то $t_\beta(9) = 1.83$. $\varepsilon_\beta = 1.83 \cdot 2.51 \approx 4.6$, то есть $m = 172.4 \pm 4.6$ с вероятностью 0.95.

Замечание

Физики обычно предпочитают приводить результат \pm одно среднее квадратичное отклонение. Из таблиц распределения Стьюдента следует, что доверительный интервал полушириной в одно среднее квадратичное отклонение содержит ис-

тинный результат с вероятностью всего 0.66.

Доверительный интервал для дисперсии

\hat{D} – сумма n случайных величин вида: $\frac{(x_i - \hat{m})}{n-1}$. Они зависимы, так как \hat{m} зависит от всех X_i . Но при $n \rightarrow \infty$ распределение их суммы приближается к нормальному. При $n \in (20, 30)$ его уже приближенно можно считать нормальным приближенным способом построения доверительного интервала. Найдем характеристики этого приближенного нормального распределения. $M(\hat{D}) = D$. (оценка \hat{D} несмещенна.) $D(\hat{D}) = \frac{1}{n}\mu_4 - \frac{n-3}{n(n-1)}D^2$

Вместо D можно взять \tilde{D} . Брать вместо μ_4 его оценку $\frac{1}{n} \sum_{i=1}^n (x_i - \hat{m})^4$ плохо, так как при небольшом n ошибка будет велика. Если известен закон распределения X , то μ_4 можно выразить через D . Для нормального распределения $X : \mu_4 = 3D^2$, следовательно, $D(\hat{D}) = \frac{3}{n}D^2 - \frac{n-3}{n(n-1)}D^2 = \frac{2}{n-1}D^2 \approx \frac{2}{n-1}\hat{D}^2$. $\sigma_{\hat{D}} = \sqrt{\frac{2}{n-1}}\hat{D}$. Пусть доверительная вероятность β . Найдем t_β : $\Phi^*(t_\beta) = \frac{\beta+1}{2}$, по таблице доверительного интервала для дисперсии: $(\hat{D} - t_\beta \sigma_{\hat{D}}, \hat{D} + t_\beta \sigma_{\hat{D}})$

Точное построение доверительного интервала для дисперсии

Известно, что $V = \frac{(n-1)\hat{D}}{D}$ имеет распределение χ^2 с $(n-1)$ степеней свободы, плотность которой:

$$K_{m-1}(U) = \begin{cases} 0, U < 0 \\ \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} U^{\frac{n-1}{2}-1} e^{-\frac{U}{2}}, U > 0 \end{cases}$$

Будем выбирать интервал l_β так, чтобы вероятности выхода V за пределы l_β влево и вправо были одинаковы и равны.

$\frac{\alpha}{2} = \frac{1-\beta}{2}$. Распределение χ^2 затабулировано. Здесь приведем числа χ^2 такие, что $P(V > \chi^2) = p$, r – число степеней свободы. ($r = n-1$).

Зафиксируем $r = n-1$ и найдем по таблице два значения χ^2 , одно соответствует $p_1 = \frac{\alpha}{2}$, другое $p_2 = 1 - \frac{\alpha}{2}$. Обозначим их χ_1^2 и χ_2^2 . Интервал $l_\beta = (\chi_2^2, \chi_1^2)$.

Доверительный интервал для дисперсии $P(D_1 < D < D_2) = \beta$ – он будет тогда, когда $V = \frac{(n-1)\hat{D}}{D}$ попадет в l_β .

Если $V < \chi_1^2$, то $\frac{(n-1)\hat{D}}{D} M < \chi_1^2$ следовательно $\frac{(n-1)\hat{D}}{\chi_1^2} < D$.

Если $V > \chi_2^2$, то $\frac{(n-1)\hat{D}}{D} M > \chi_2^2$ следовательно $\frac{(n-1)\hat{D}}{\chi_2^2} > D$.

То есть доверительный интервал для дисперсии: $(\frac{(n-1)\hat{D}}{\chi_1^2}, \frac{(n-1)\hat{D}}{\chi_2^2})$

7.4 Оценки числовых характеристик случайных величин

Система (X, Y) . Результат n опытов $(x_1, y_1), \dots, (x_n, y_n)$.

Несмещенные оценки:

$$\hat{m}_x = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{m}_y = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$\hat{D}_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m}_x)^2, \quad \hat{D}_y = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{m}_y)^2.$$

$$\hat{K}_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m}_x)(y_i - \hat{m}_y)$$

7.5 Проверка гипотез

Рассмотрим задачу проверки опытных данных заданому закону распределения. (Критерия согласованности χ^2 Пирсона по измерению X .)

Пусть в результате опытов (независимых) получены результаты, которые сведены в статистический ряд. p_i^* – частоты.

Мы предполагаем, что X распределена по некоторому "теоретическому" закону. На основании его вычисляем теоретические вероятности попадания случайной

величины в $(x_1, x_2), \dots, (x_k, x_{k+1})$. Мера расхождения $U = \sum_{i=1}^k c_i (p_i^* - p_i)^2$. "Весо-

вые" коэффициенты надо выбирать обратно пропорционально вероятностям, так как одно и то же значение $p^* - p$ более значимо если само p_i мало. Возьмем $c_i = \frac{n}{p_i}$. Тогда U будет распределено по "закону χ^2 ".

Замечание

Пусть X_i – независимые случайные величины, распределенные, по нормальному закону с $M(X_i) = 0$, $\sigma_{x_i} = 1$. Тогда величина $U = \sum_{i=1}^r x_i^2$ распределена по закону

$$\text{с } \chi^2 \text{ с } r \text{ степенями свободы. У нас: } U = \chi^2 = \sum_{i=1}^k \frac{n(p_i^* - p_i)^2}{p_i} = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i},$$

так как $p_i^* = \frac{m_i}{n}$.

Число степеней свободы r равно числу разрядов k минус число независимых условий, наложенных на частоты p_i^* . Одно из таких условий, которое всегда на-

кладывается : $\sum_{i=1}^k p_i^* = 1$

Если все параметры "теоретического" распределения заранее известны, то боль-

ше условий нет. Если эти параметры мы оцениваем из результатов опытов, то добавляется еще условие, например $\sum_{i=1}^k x_i^* p_i^* = m_x$, $\sum_{i=1}^k (x_i^* - \hat{m}_x)^2 p_i^* = D_x$, и соответственно уменьшается число степеней свободы. Рассмотрим на примере действие критерия χ^2 .

Пример

20 лет собирали сведения о количестве офицеров русской армии, погибших в результате гибели под ними коня. (на основании ежегодных докладов 10 армейских корпусов, то есть 200 донесений.

Пусть X – число погибших в одном корпусе за год.

‘ Полное число погибших: $0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1 = 122$.

Среднее число погибших: $\frac{122}{200} = 0.61$. Проверим гипотезу о том, что эти данные описывают распределение Пуассона: $P(x = x_i) = p_i = \frac{\mu^i e^{-\mu}}{i!}$. Для определения p_i надо оценить одним параметром μ . В качестве μ возьмем $\hat{\mu} = 0.61$. np_i приведен в третьей строке. Так как ожидаемые частоты для $x_i > 2$ малы, объединим их $U = \chi^2 = \sum \frac{(m_i - np_i)^2}{np_i} = \frac{(0.3)^2}{108.7} + \frac{(1.3)^2}{66.3} + \frac{(1.8)^2}{20.2} + \frac{(0.8)^2}{4.8} = 0.32$. Есть 2 линей-

ки соотношений, связывающие частоты: $\sum m_i = 200$ и $\sum x_i m_i = 122 = \hat{\mu}$.

То есть число степеней свободы : $r = 4 - 1 - 1 = 2$. То есть U распределено примерно по закону χ^2 с двумя степенями свободы. U распределено примерно по закону χ^2 с двумя степенями свободы.

Из таблицы распределения χ^2 с $r = 2$ находим, что с вероятностью 0.95 χ^2 должно лежать в интервале от 0 до 6.0 так как наше значение 0.32 попадает в этот незначимый интервал, то проверяемая гипотеза не отвергается.

В таблице указана вероятность p того, что $\chi^2 \geq \chi_p^2$ при числе степеней свободы r .

Если бы мы получили для χ^2 значение 6.1 или больше, то должны были бы считать его значимым, так как если проверяемая гипотеза верна, то такие значения χ^2 может принимать с вероятностью не более 0.05. 0.05 называется уровнем значимости.

Обычно, схема действий такая: По экспериментальным данным составляется функция U и на основе "нулевой гипотезы" H вычисляется её распределение. Затем проверяют, как изменится U , если гипотезы H неверна (верна \bar{H}).

Обычно U в этом случае намного больше. На рисунке показано два распределения величины U , соответствующие двум конкурирующим гипотезам $H(a)$ и $\hat{H}(\hat{a})$. Если из опыта получилось большое значение U , то можно сказать, то либо справедлива гипотеза H , но произошло чрезвычайно маловероятное событие, либо

более правдоподобна гипотеза \overline{H} . Если оставаться в рамках одной гипотезы, то результат следует признать значимым в том смысле, что он выходит за рамки обычных статистических ошибок.

Определение

Отвергая нулевую гипотезу H , когда она справедлива, мы совершаем ошибку первого рода. Принимая нулевую гипотезу H , когда справедлива конкурирующая гипотеза \overline{H} , мы совершаем ошибку второго рода. Вероятность ошибки первого рода называется уровнем значимости (α). Если вероятность ошибки второго рода равна β , то $1 - \beta$ называют мощностью критерия.

Часто возможен компромисс между уровнем значимости и мощностью критерия. Если $U > U_\alpha$, мы отвергаем H и принимаем \overline{H} , если $U < U_\alpha$ – наоборот.

7.6 Метод моментов

Пусть функция распределения случайной величины X зависит от k параметров $a_1, a_2 \dots a_k$. Надо выбрать оценки этих параметров по результатам n опытов. Теоретические моменты $(M(X), D(X), \dots \mu_k(x))$ являются функциями от этих параметров. По результатам опытов находим оценки моментов $\hat{M}(X), \hat{D}(X), \dots \hat{\mu}_k(x)$. Оценки параметров $a_1, a_2 \dots a_k$ находим решая систему уравнений:

$$\begin{cases} m_x = \hat{m}_x \\ D(X)(a_1, \dots a_k) = \hat{D}(X) \\ \dots \\ \mu_k(x)(a_1, \dots a_k) = \hat{\mu}_k(X) \end{cases}$$

Пример

Найти точечные оценки параметров a, b равномерно распределенных в интервале (a, b) по результатам n опытов. По результатам n опытов находим $\hat{m}_x = \bar{x}$, и $\hat{D}(X) = \hat{\sigma}_x^2$.

Теоретические моменты:

$$M(X) = \frac{a+b}{2} = \bar{X}$$

$$D(X) = \frac{(b-a)^2}{12} = \hat{D}_x = \hat{\sigma}_x^2.$$

Решая эту систему, находим оценки для a и b : $\hat{a} = \bar{X} - \hat{\sigma}_x \sqrt{3}$, $\hat{b} = \bar{X} + \hat{\sigma}_x \sqrt{3}$.

Замечание

Об ошибке прибора при измерении. Если в паспорте прибора задан гарантированный его производителем интервал $\pm \Delta$, из которого не выходят предельные отклонения (ошибки). Тогда дисперсия ошибки $D = \frac{\Delta^2}{3}$. Если задан класс точности прибора (a), то $\Delta = \frac{a \cdot b}{100}$, где b – максимальное значение шкалы прибора.

7.7 Принцип максимального правдоподобия

Пусть есть урна, в которой неизвестная смесь черных и белых шаров. Вероятность вытянуть белый шар – p , может принимать любое значение от 0 до 1.

Хотим изменить p , вынимая шар из урны. Пусть мы вытащили шар n раз (с возвращением) и белый шар оказался в r случаях. Вероятность этого события $l(p) = C_n^r p^r (1-p)^{n-r} = l(p|r)$ эта функция при известном r (и n) и неизвестном p называется функцией правдоподобия. Выбираем то значение p , которое с наибольшей вероятностью приводит к наблюдаемому значению r . Когда параметр p дискретен, надо вычислить для разных p функцию правдоподобия и выбрать то p , при котором функция правдоподобия максимальна. Если p непрерывен, то можно искать максимум с помощью дифференцирования: $\frac{dl}{dp} = C_n^r (r p^{r-1} (1-p)^{n-r} - (n-r) p^r (1-p)^{n-r-1}) = 0$
 $r - rp - np - rp = 0, p = \frac{r}{n}$.

Определение

Логарифмическая функция правдоподобия:

$L(p|r) = \ln l(p|r) = r \ln p + (n-r) \ln(1-p)$. (здесь отброшено постоянное слагаемое, а в $l(a|r)$ – постоянный множитель, который не влияет на максимум.)

$L(p|r)$ имеет максимум при том же значении p , что и $l(p|r)$.

Более общий случай:

страница 62 Пусть X – дискретная случайная величина со значениями $x_1 \dots x_l$ и

$P(X = x_j; \Theta) = p_j(\Theta)$. $\Theta = (\Theta_1, \dots, \Theta_k)$ – неизвестные параметры. Пусть m_1, \dots, m_l – число появлений значений x_1, \dots, x_l в выборке X_1, X_2, \dots, X_n . Тогда функция правдоподобия значений $l(\Theta|x_1, x_2, \dots, x_n) = p_1^{m_1}(\Theta) \cdot p_2^{m_2}(\Theta) \cdot \dots \cdot p_l^{m_l}(\Theta)$.
 $L(\Theta|x_1, x_2, \dots, x_n) = \ln l(\Theta|x_1, x_2, \dots, x_n)$

Пусть случайная величина X имеет непрерывное распределение с плотностью $f(x|\Theta)$, Θ – единственный параметр. Плотность вероятности выборки из n измерений (n независимых случайных величин):

$$\prod_{i=1}^n f(x_i|\Theta) = l(\Theta|x_1, x_2, \dots, x_n)$$

Логарифмическая функция правдоподобия: $L(\Theta|x_1, \dots, x_n) = \sum_{i=1}^n \ln f(x_i|\Theta)$

Если параметров несколько, то $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_k)$. Параметры выбираются из

условия максимума функции l (или L)

Пример

$$f(x|m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$l(\mu, \sigma^2 | x_1, \dots, x_n) = (\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}}$$

$$L(\mu, \sigma^2 | x_1, \dots, x_n) = -\frac{n}{2} \ln \sigma^2 - \sum_{i=1}^n (x_i - m)^2 \frac{1}{2\sigma^2}$$

$$\frac{\delta L}{\delta m} = 0, \frac{\delta L}{\delta \sigma^2} = 0$$

Отсюда находим $\hat{m}, \hat{\sigma}^2$

$$\frac{\Delta L}{\Delta m} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \hat{m}) = 0, \text{ следовательно } \hat{m} = \bar{x}$$

$$\frac{\Delta L}{\Delta \sigma^2} = -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - m)^2 = 0, \text{ следовательно } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

То есть тут оказывается смещенная оценка дисперсии.