

Министерство образования Российской Федерации

*“МАТИ”- РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ им. К. Э. ЦИОЛКОВСКОГО*

Кафедра ”Вышая математика”

А. С. Кочуров

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

(краткий конспект)

Москва 2003 г.

Лекция 1.

Предмет теории вероятностей. Случайные события. Алгебра событий. Относительная частота и вероятность случайного события. Классическое определение вероятности. Основные свойства вероятности. Основные формулы комбинаторики.

Литература.

Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М., Высшая школа, 1998.

Гмурман В.Е. Теория вероятностей и математическая статистика. М., Высшая школа, 1977.

Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика. М., Высшая школа, 1982

Чистяков В.П. Курс теории вероятностей. М., Наука, 1988.

Предмет теории вероятностей. Случайные события.

Испытание (наблюдение) – наличие определенного комплекса условий.

Возможный результат – исход испытания или наблюдения – называется событием.

Пример: Бросание монеты (испытание). Результат испытания – событие: 1. выпадение герба, 2. выпадение решетки.

Некоторые из событий принято называть достоверными (в данном испытании), т.е. те, которые при данном испытании неизбежно реализуются. Если же некоторое событие (в данном испытании) не может произойти, то его называют невозможным.

Определение. Результат испытания, который нельзя заранее прогнозировать, называется случайным событием. Теория вероятностей изучает закономерности случайных событий.

Алгебра событий.

С каждым испытанием связан ряд интересующих нас событий, которые, вообще говоря, могут появляться одновременно. Так, при бросании игральной кости, событиями будут:

1. Выпадение 1 очка (событие A),
2. Выпадение нечетного числа очков (событие B). Понятно, что эти события могут происходить одновременно. В этом случае говорят, что события A и B не исключают друг друга.

Обычно с каждым испытанием связывают полную систему элементарных исходов (или элементарных событий), взаимно исключающих друг друга. Это означает, что

- 1). Каждый исход испытания представляется одним и только одним элементарным событием,
- 2). Каждое событие A , связанное с рассматриваемым испытанием, есть множество, состоящее из конечного или бесконечного числа элементарных событий рассмотренной системы,
- 3). Событие A происходит тогда и только тогда, когда реализуется одно из элементарных событий, входящих в это множество.

Пример. Событие A состоит в выпадении нечетного числа очков при однократном бросании игральной кости. За элементарные события могут быть приняты следующие результаты испытания: выпадение (1), или (2), или (3), или (4), или (5), или (6). При этом событие A станет множеством $\{(1), (3), (5)\}$.

Таким образом, с каждым испытанием связывают систему элементарных событий $\Omega = \{w\}$.

Однако элементарные события не исчерпывают полный набор событий рассматриваемого испытания. Полный набор событий представляет собой некоторую систему $\mathcal{A}$ подмножеств Ω . Принято считать, что эта система является σ -алгеброй, саму же систему $\mathcal{A}$ называют σ -алгеброй событий. Это означает, что

1. Эта система содержит пустое множество $\emptyset$ и множество Ω : $\emptyset \in \mathcal{A}$, $\Omega \in \mathcal{A}$,
2. Вместе с каждым множеством $A \in \mathcal{A}$ множество $\bar{A} \in \mathcal{A}$,
3. Если $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$, то и $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

При определении σ -алгебры широко используются хорошо известные операции над множествами: 1. объединение $A \cup B$, 2. пересечение $A \cap B$, 3. дополнение $\bar{A}$ до Ω . Кроме того предполагается известными свойства этих операций

1. коммутативность объединения и пересечения множеств,
2. ассоциативность объединения и пересечения множеств,
3. дистрибутивность этих операций относительно друг друга,
4. $A \cup \emptyset = A$, $A \cap \Omega = A$,

5. $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$,
6. $A \cup A = A$, $A \cap A = A$,
7. $A \cup \Omega = \Omega$, $A \cap \emptyset = \emptyset$,
8. законы де Моргана $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$,
9. законы поглощения $A \cup (A \cap B) = A$, $A \cap (A \cup B) = A$.

Часто используются также знаки принадлежит $\in$, является подмножеством $\subset$, совпадает $=$.

Чтобы определить вероятность необходимо указать правило (функцию вероятности), сопоставляющее каждому $A \in \mathcal{A}$ число $P(A) \in [0, 1]$, удовлетворяющее определенным аксиомам вероятности. Поясним эти аксиомы на примере классического определения вероятности.

Относительная частота и вероятность случайного события. Классическое определение вероятности. Основные свойства вероятности.

Пусть некоторое испытание имеет лишь конечное число элементарных исходов (событий):

$$\Omega = \{w_1, w_2, \dots, w_n\}.$$

В рассматриваемой модели в качестве $\mathcal{A}$ выбирают всевозможные подмножества Ω . Кроме того, по определению, считают, что все элементарные исходы равновероятны, т.е. $P(w_i) = 1/n$, $i = 1, \dots, n$ и более общо

$$P(A) = |A|/n$$

для каждого $A \subset \Omega$ (здесь $|A|$ — число элементов в A).

Легко проверить следующие свойства так определенной функции вероятности

1. $P(\emptyset) = 0$ (вероятность невозможного события равна нулю),
2. $P(\Omega) = 1$ (вероятность достоверного события равна единице),
3. $P(A) \geq 0$ для всех $A \in \mathcal{A}$,
4. $P(A + B) = P(A) + P(B)$, если $A \cap B = \emptyset$,
5. $P(A) \leq P(B)$, если $A \subset B$,
6. $P(A + \bar{A}) = P(A) + P(\bar{A}) = 1$.

Свойства 2)-4) являются основополагающими при определении функции вероятности в общем случае. Именно их и принимают за аксиомы, которым удовлетворяет эта функция. Только условие 4 заменяют похожим, но несколько более сильным условием:

$$4'. P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n), \text{ если } A_i \cap A_j = \emptyset \text{ при } i \neq j.$$

Пример: Игральная кость бросается 2 раза. Какова вероятность того, что сумма выпавших очков равна 6 (событие A)?

Равновозможные элементарные исходы: (x, y) , где x и y принимают значения 1, 2, 3, 4, 5, 6. Общее число элементарных исходов равно 36. Событию A благоприятствуют 5 пар (1,5), (2,4), (3,3), (4,2), (5,1). Следовательно, $P(A) = 5/36$.

Классическое определение вероятности предполагает, что 1) число элементарных исходов конечно, 2) эти исходы равновозможны.

Однако на практике встречаются испытания с бесконечным числом различных возможных исходов. Кроме того нет общих методов, позволяющих результат испытания представить в виде суммы равновозможных элементарных исходов. Поэтому применяют иное, статистическое определение вероятности. Пусть производится n однотипных испытаний, одним из исходов которых является данное событие A .

Определение. Отношение числа появлений m события A к общему числу испытаний n называется относительной частотой $W_n(A)$ события A .

При однотипных массовых испытаниях во многих случаях наблюдается устойчивость относительной частоты события, т.е. при числе испытаний $n \rightarrow \infty$ относительная частота $W_n(A)$ события A стремится к некоторой постоянной p . Это число p и называется вероятностью события A в статистическом смысле.

Пример: В результате ряда испытаний было обнаружено, что при 200 выстрелах стрелок попадает в цель в среднем 190 раз. Какова вероятность p поражения цели этим стрелком? Сколько для него попаданий в цель можно ожидать при 1000 выстрелах?

Согласно статистическому определению вероятности $p = 190/200 = 95\%$. Поэтому число удачных выстрелов должно примерно составлять $1000 \cdot p = 950$.

Основные формулы комбинаторики.

Рассмотрим совокупность n различных элементов $a_1, a_2, \dots, a_n$. Произвольную упорядоченную выборку (возможно с повторениями) размера m из этих элементов $a_{i_1}, a_{i_2}, \dots, a_{i_m}$ будем называть соединением ($1 \leq i_k \leq n, k = 1, \dots, m$). Легко подсчитать, что число различных соединений равно n^m .

Размещениями из n элементов по m ($m \leq n$) называются такие их соединения, каждое из которых содержит ровно m различных элементов (выбранных из данных элементов) и которые отличаются или самими элементами или порядком их следования. Нетрудно подсчитать, что общее число размещений A_n^m из n элементов по m равно

$$A_n^m = \frac{n!}{(n-m)!},$$

где $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$ и, по определению $0! = 1$.

Перестановками из n элементов называются такие их соединения, каждое из которых содержит все n исходных элементов и которые отличаются порядком следования этих элементов. Таким образом, перестановки – это размещения из n элементов по n . Общее число различных перестановок равно $A_n^n = \frac{n!}{(0)!} = n!$.

Сочетаниями из n элементов по m ($m \leq n$) называются такие их соединения, каждое из которых содержит ровно m различных элементов (выбранных из данных элементов) и каждые два из которых отличаются хотя бы одним элементом. Нетрудно понять, что осуществляя все $m!$ перестановок в каждом из сочетаний, мы получим все размещения. Таким образом, для общего числа C_n^m сочетаний из n элементов по m имеем формулу

$$C_n^m \cdot m! = A_n^m, \quad C_n^m = \frac{n!}{(n-m)!m!}.$$

Известно, что числа C_n^m являются коэффициентами в формуле бинома Ньютона

$$(p+q)^n = p^n + C_n^1 p^{n-1} q + C_n^2 p^{n-2} q^2 + \dots + C_n^{n-1} p q^{n-1} + q^n.$$

Эти числа, в частности, обладают следующими простыми свойствами $C_n^m = C_n^{n-m}$ и

$$C_n^m + C_n^{m+1} = \frac{n!}{(n-m)!m!} + \frac{n!}{(n-m-1)!(m+1)!} = \frac{n!}{(n-m-1)!m!} \left(\frac{1}{n-m} + \frac{1}{m+1} \right) = C_{n+1}^{m+1}.$$

Лекция 2.

Геометрические вероятности. Теорема сложения вероятностей. Противоположные события. Условные вероятности. Теорема умножения вероятностей. Независимые события. Формула полной вероятности.

Рассмотрим второй класс вероятностных пространств $(\Omega, \mathcal{A}, P)$. Пусть Ω – некоторое множество на прямой, плоскости, в пространстве $\mathbb{R}^3$ или в пространстве большего числа измерений. Пусть $V(A)$ – длина, площадь или объем (в зависимости от расположения Ω) множества $A \subset \Omega$. Для сокращения $V(A)$ называют объемом даже если это множество на прямой или плоскости. Пусть $V(\Omega) < \infty$, $\mathcal{A}$ – совокупность подмножеств $A \subset \Omega$, для которых можно вычислить их объем $V(A)$. Тогда можно показать, что $\mathcal{A}$ – алгебра множеств. Определим функцию вероятности на этой алгебре множеств правилом

$$P(A) = \frac{V(A)}{V(\Omega)}.$$

Пользуясь свойствами кратного интеграла нетрудно проверить выполнение аксиом функции вероятности

1. $P(\Omega) = V(\Omega)/V(\Omega) = 1$ (вероятность достоверного события равна единице),
2. $P(A) = V(A)/V(\Omega) \geq 0$ для всех $A \in \mathcal{A}$,
3. $P(A+B) = V(A+B)/V(\Omega) = (V(A) + V(B))/V(\Omega) = P(A) + P(B)$, если $A \cap B = \emptyset$.

Рассмотренная модель введения вероятности называется геометрическим определением вероятности. Элементарными исходами в этой модели являются точки множества Ω и вероятность каждого элементарного исхода равна нулю (так как объем любой точки равен нулю). Положительную вероятность имеют

лишь большие соединения элементарных исходов (большие множества). Этот пример введения вероятности показывает, что бывают события, которые могут происходить (т.е. они не являются невозможными), но вероятность наступления которых равна нулю.

Рассмотрим некоторые следствия аксиом 1)-3). Предположим, что A и B – произвольные события из $\mathcal{A}$. Обозначим $C_1 = A \cap B$, $C_2 = A \setminus C_1$, $C_3 = B \setminus C_1$. Тогда $A = C_1 \cup C_2$, $B = C_1 \cup C_3$,

$$P(A \cup B) = P(C_1 \cup C_2 \cup C_3) = P(C_1) + P(C_2) + P(C_3), \quad (1)$$

поскольку события C_1, C_2, C_3 взаимоисключающие: $C_1 \cap C_2 = \emptyset$, $C_1 \cap C_3 = \emptyset$, $C_2 \cap C_3 = \emptyset$. С другой стороны

$$P(A) = P(C_1 \cup C_2) = P(C_1) + P(C_2), \quad P(B) = P(C_1 \cup C_3) = P(C_1) + P(C_3). \quad (2)$$

Совмещая формулы (1) и (2), получим

$$P(A \cup B) = P(A) + P(B) - P(C_1) = P(A) + P(B) - P(A \cap B).$$

Наряду с аксиомой 3) полученная формула также называется формулой для вероятности суммы двух событий. В отличие от 3), она справедлива для любых событий A и B . Ее можно распространить на сумму большего числа событий:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

и т.д.

Полная группа событий.

Другим следствием аксиом 1)-3) является уже рассматривавшаяся связь между событием A и противоположным к нему событием $\bar{A} = \Omega \setminus A$. Поскольку события A и $\bar{A}$ взаимоисключающие, то

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

Система событий $A_1, A_2, \dots, A_n$ называется полной группой событий для данного испытания, если любым исходом его является одно и только одно событие этой группы. Иными словами для полной группы событий выполняются следующие условия:

- 1) $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$ – сумма событий достоверна,
- 2) события A_i, A_j несовместны при $i \neq j$: $A_i \cap A_j = \emptyset$.

Простейшим примером полной группы событий является пара событий A и $\bar{A}$. Также как и для пары A и $\bar{A}$, для полной группы событий сумма вероятностей входящих в группу событий равна 1:

$$1 = P(\Omega) = P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

Условные вероятности. Теорема умножения вероятностей. Независимые события.

Пусть $(\Omega, \mathcal{A}, P)$ – какое-нибудь вероятностное пространство, $B \in \mathcal{A}$, $P(B) > 0$. Образует новое вероятностное пространство $(B, \mathcal{B}, P_B)$. В нем алгебра множеств $\mathcal{B}$ образована всевозможными пересечениями $A \cap B$ множеств $A \in \mathcal{A}$ и множества B , а функция P_B определяется равенствами

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

Проверим выполнение аксиом алгебры множеств:

- 1) Так как $\emptyset \in \mathcal{A}$, $B \in \mathcal{A}$, то $\emptyset = \emptyset \cap B \in \mathcal{B}$, $B = B \cap B \in \mathcal{B}$,
- 2) Пусть $A \in \mathcal{A}$ и, значит, по определению $A \cap B \in \mathcal{B}$. Тогда дополнение $A \cap B$ в пространстве B имеет вид $B \setminus A$. Для него

$$B \setminus A = B \cap \bar{A},$$

так как

$$\Omega = A \cup \bar{A}, \quad A \cap \bar{A} = \emptyset, \quad B = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A})$$

и $(B \cap A), (B \cap \bar{A})$ не имеют общих точек. Поскольку $\bar{A} \in \mathcal{A}$, отсюда следует, что $B \setminus A \in \mathcal{B}$.

3) Пусть $A, C \in \mathcal{A}$ и, значит, по определению $A \cap B$, $C \cap B$ принадлежат $\mathcal{B}$. Так как $A \cup C \in \mathcal{A}$, то, пользуясь дистрибутивным законом, связывающим операции объединения и пересечения множеств получим

$$(A \cup C) \cap B = (A \cap B) \cup (C \cap B) \in \mathcal{B},$$

что и требовалось проверить.

Теперь проверим выполнение аксиом функции вероятности:

$$1) P_B(A \cap B) = \frac{P(A \cap B)}{P(B)} \geq 0,$$

$$2) P_B(B \cap B) = \frac{P(B \cap B)}{P(B)} = 1,$$

$$3) P_B((A \cap B) \cup (C \cap B)) = \frac{P((A \cap B) \cup (C \cap B))}{P(B)} = \frac{P(A \cap B) + P(C \cap B)}{P(B)} = P_B(A \cap B) + P_B(C \cap B),$$

если события $A \cap B$, $C \cap B$ – взаимоисключающие в B (т.е. $(A \cap B) \cap (C \cap B) = \emptyset$).

Так определенная модель вероятностного пространства называется еще условной вероятностью. Для нее принято специальное обозначение

$$P(A | B) = P_B(A \cap B)$$

– вероятность наступления события A при условии, что произошло событие B (условная вероятность события A). При этом вероятность $P(A)$ события A иногда называют безусловной вероятностью. Еще раз подчеркнем обязательное условие определения условной вероятности: $P(B) > 0$.

Пример. В урне находятся 7 белых и 3 черных шара. Какова вероятность:

- 1) извлечения из урны белого шара (событие A), $P(A) = 0.7$,
- 2) извлечения из урны белого шара после удаления белого шара (событие B), $P(A | B) = 0.66 \dots$,
- 3) извлечения из урны белого шара после удаления черного шара (событие C), $P(A | C) = 0.77 \dots$

Два события A , B , для которых $0 < P(A), P(B) < 1$, называются независимыми, если вероятность каждого из них не зависит от появления или не появления другого, т.е.

$$P(A) = P(A | B) = P(A | \bar{B}), \quad P(B) = P(B | A) = P(B | \bar{A}).$$

Если хотя бы одно из этих равенств не выполняется, то события называются зависимыми. Из определения следует, что для независимых событий выполняется равенство

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

и, значит, $P(A \cap B) = P(A)P(B)$. По аналогичной причине выполняются еще три равенства

$$P(A \cap \bar{B}) = P(A)P(\bar{B}), \quad P(\bar{A} \cap B) = P(\bar{A})P(B), \quad P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B}).$$

Выполнение последних 4 равенств иногда принимают за определение независимости событий A и B . Про эти равенства также говорят, что вероятность совместного появления двух независимых событий равна произведению вероятностей этих событий.

Пример. Вероятность поражения цели первым стрелком равна 0.9 (событие A), вероятность поражения цели вторым стрелком равна 0.8 (событие B). Какова вероятность поражения цели хотя бы одним стрелком (событие C)?

$$P(\bar{C}) = P(\bar{A}\bar{B}) = 0.1 \cdot 0.2 = 0.02, \quad P(C) = 0.98.$$

Определение независимости распространяется и на большее число событий. Несколько событий называются независимыми в совокупности, если любое из них и любое пересечение остальных (включающее либо часть из них, либо все оставшиеся) есть события независимые. Для независимых событий вероятность одновременного наступления событий равна произведению вероятностей этих событий:

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Пусть A_1, A_2, A_3, A_4 – некоторые события вероятностного пространства $(\Omega, \mathcal{A}, P)$. Следующая формула (и ее естественное обобщение на большее число событий) иногда называется теоремой умножения

$$\begin{aligned} P(A_1 A_2 A_3 A_4) &= P(A_1) \frac{P(A_1 A_2)}{P(A_1)} \frac{P(A_1 A_2 A_3)}{P(A_1 A_2)} \frac{P(A_1 A_2 A_3 A_4)}{P(A_1 A_2 A_3)} = \\ &= P(A_1) P(A_2 | A_1) P(A_3 | A_1 A_2) P(A_4 | A_1 A_2 A_3) \end{aligned}$$

(при ее написании предполагается, что все входящие в нее события имеют ненулевые вероятности).

Формула полной вероятности.

Пусть $(\Omega, \mathcal{A}, P)$ – вероятностное пространство, $\{H_1, H_2, \dots, H_n\}$ – полная группа несовместных событий (назовем эти события гипотезами) этого пространства, $P(H_i) > 0$, $i = 1, \dots, n$, пусть A – произвольное событие. Тогда

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

Эта формула и называется формулой полной вероятности. Проверим ее:

$$A = A \cap \Omega = A \cap (H_1 \cup H_2 \cup \dots \cup H_n) = (A \cap H_1) \cup (A \cap H_2) \cup \dots \cup (A \cap H_n).$$

Поэтому, по определению условной вероятности,

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

Пример. В магазин для продажи поступает продукция 3 фабрик, относительные доли которых составляют соответственно 50, 30 и 20 процентов. Для продукции фабрик брак составляет соответственно 2, 3 и 5 процентов. Какова вероятность покупки в магазине доброкачественного продукта (событие A)?

H_1, H_2, H_3 – гипотезы приобретения продукта, произведенного на 1-й, 2-й и 3-й фабриках ($P(H_1) = 0.5$, $P(H_2) = 0.3$, $P(H_3) = 0.2$). По формуле полной вероятности

$$P(A) = P(H_1)P(A | H_1) + P(H_2)P(A | H_2) + P(H_3)P(A | H_3) = 0.5 \cdot 0.98 + 0.3 \cdot 0.97 + 0.2 \cdot 0.95 = 0.971.$$

Лекция 3.

Формула Бейеса. Схема и формула Бернулли. Теоремы Пуассона и Муавра-Лапласа.

Пусть $\{H_1, \dots, H_n\}$ – полная группа несовместных гипотез вероятностного пространства $(\Omega, \mathcal{A}, P)$, $P(H_i) > 0$, $i = 1, \dots, n$, – вероятности этих гипотез. Производится опыт, в результате которого зафиксировано появление события $A \in \mathcal{A}$, $P(A) > 0$, причем известно, что этому событию гипотезы $\{H_1, \dots, H_n\}$ приписывали вероятности $P(A | H_i)$, $i = 1, \dots, n$. Спрашивается, каковы будут вероятности этих гипотез после опыта (вероятности апостериори). Иными словами требуется определить условные вероятности

$$P(H_i | A), \quad i = 1, \dots, n,$$

(переоценить вероятности исходных гипотез). На основании теоремы умножения вероятностей

$$P(A \cap H_i) = P(H_i) \cdot P(A | H_i) = P(A) \cdot P(H_i | A), \quad P(H_i | A) = \frac{P(H_i) \cdot P(A | H_i)}{P(A)}.$$

Вероятность события $P(A) > 0$ при условии, что заданы величины $P(H_i)$, $P(A | H_i)$, $i = 1, \dots, n$, подсчитывается по формуле полной вероятности

$$P(A) = \sum_{j=1}^n P(H_j) \cdot P(A | H_j).$$

Отсюда имеем формулу вероятностей гипотез после опыта (формулу Байеса)

$$P(H_i | A) = \frac{P(H_i) \cdot P(A | H_i)}{\sum_{j=1}^n P(H_j) \cdot P(A | H_j)}, \quad i = 1, \dots, n.$$

Пример. Вероятность поражения самолета при одиночном выстреле для первого ракетного расчета (событие A) равна 0.2, а для 2-го (событие B) – 0.1. Каждое из орудий производит 1 выстрел, причем зафиксировано одно попадание в самолет (событие C). Какова вероятность того, что удачный выстрел принадлежит первому расчету?

До опыта возможны 4 гипотезы $H_1 = A \cap B$, $H_2 = A \cap \bar{B}$, $H_3 = \bar{A} \cap B$, $H_4 = \bar{A} \cap \bar{B}$. Вероятности этих гипотез при независимом действии расчетов равны $P(H_1) = 0.2 \cdot 0.1 = 0.02$, $P(H_2) = 0.18$, $P(H_3) = 0.08$, $P(H_4) = 0.72$. Если произошло в точности одно попадание, то гипотезы H_1 , H_4 невозможны:

$$P(C | H_1) = P(C | H_4) = 0, \quad P(C | H_2) = 1, \quad P(C | H_3) = 1.$$

Следовательно гипотезы H_1 , H_4 отпадают, а вероятности гипотез H_2 , H_3 подсчитываются по формуле Байеса

$$P(H_2 | C) = \frac{0.18 \cdot 1}{0.18 \cdot 1 + 0.08 \cdot 1} \approx 0.7, \quad P(H_3 | C) = \frac{0.08 \cdot 1}{0.18 \cdot 1 + 0.08 \cdot 1} \approx 0.3.$$

Биномиальный закон распределения вероятностей. Схема и формула Бернулли.

Рассмотрим два каких-либо испытания: $(\Omega_1, \mathcal{A}_1, P_1)$, $(\Omega_2, \mathcal{A}_2, P_2)$. Пусть w_1 – элементарное событие первого пространства (Ω_1) , w_2 – элементарное событие второго пространства (Ω_2) . Тогда пару событий (w_1, w_2) можно рассматривать как элементарное событие пространства $(\Omega_1 \times \Omega_2)$. Не будем определять как выглядят алгебра событий и функция вероятности в новом пространстве в общем случае, а скажем, что в случае, когда исходные пространства представляли классическую вероятностную модель, то и новое пространство естественно считать таким же. В нем алгебра событий – совокупность всех подмножеств $\Omega_1 \times \Omega_2$, а вероятность подсчитывается по классическому правилу. Отметим также, что очень часто рассматриваются не пара, а большее число каких-либо испытаний и из них, по аналогичному алгоритму, конструируется сложное испытание (система испытаний).

Систему из двух испытаний будем называть независимой, если для любых событий $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$

$$P((A_1, A_2)) = P_1(A_1)P_2(A_2)$$

– вероятность наступления сложного события (A_1, A_2) равна произведению вероятностей $P_1(A_1)$ и $P_2(A_2)$ исходных пространств. Аналогичное определение независимости делается и для большего числа испытаний.

Рассмотрим теперь простейший случай системы испытаний – повторения одного и того же испытания при постоянных условиях, причем в качестве элементарных исходов каждого отдельного испытания будем различать лишь два исхода: появление некоторого события A и его непоявление $\bar{A}$.

Серия повторных независимых испытаний, в каждом из которых данное событие A имеет одну и ту же вероятность $P(A) = p$, независящую от номера испытания, называется схемой Бернулли. Таким образом в схеме Бернулли для каждого испытания имеется только два исхода: 1) событие A ("успех"), 1) событие $\bar{A}$ ("неудача") с постоянными вероятностями $P(A) = p$, $P(\bar{A}) = 1 - p = q$.

Рассмотрим задачу: в условиях схемы Бернулли определить вероятность $P_n(m)$ ($0 \leq m \leq n$) того, что при n испытаниях событие A произойдет ровно m раз. Благоприятные серии испытаний здесь имеют вид $A_1, \dots, A_n$, где A_i – это или A или $\bar{A}$, причем событие A встречается ровно m раз. Так испытания независимы, то вероятность появления каждой такой серии равна $p^m q^{n-m}$. Все такие серии получаются в результате выбора m различных номеров из общего числа n испытаний. Таким образом, всего благоприятных серий C_n^m . Так как появление различных серий одновременно не может произойти, то по формуле сложения вероятностей для случая несовместных событий получаем формулу Бернулли:

$$P_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{(n-m)!m!} p^m q^{n-m}.$$

Она называется биномиальной, так как ее правая часть представляет из себя $(m+1)$ -й член бинома Ньютона

$$(q + p)^n = q^n + C_n^1 q^{n-1} p + C_n^2 q^{n-2} p^2 + \dots + C_n^{n-1} q p^{n-1} + p^n.$$

Отсюда получаем биномиальное распределение вероятностей числа появлений события A при n независимых испытаниях

$$1 = (q + p)^n = P_n(0) + P_n(1) + P_n(2) + \dots + P_n(n-1) + P_n(n).$$

Пример. Найти вероятность того, что при 10-кратном бросании монеты герб выпадет ровно 5 раз. По формуле Бернулли

$$P_{10}(5) = \frac{10!}{5!5!} (1/2)^5 (1/2)^5 = \frac{252}{1024} \approx 0.25.$$

Если число испытаний n велико, то вычисления по формуле Бернулли становятся затруднительными. Лаплас получил важную приближенную формулу для вероятности $P_n(m)$ появления события A точно m раз, если n — достаточно большое число.

Теорема Лапласа (локальная). Пусть $p = P(A)$ — вероятность события A , $0 < p < 1$. Тогда вероятность того, что в условиях схемы Бернулли событие A при n испытаниях появится точно m раз, выражается приближенной формулой Лапласа

$$P_n(m) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(m-np)^2}{2npq}} = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{t^2}{2}}, \quad \text{где } q = 1 - p, \quad t = \frac{m - np}{\sqrt{npq}}.$$

Можно показать, что относительная погрешность приведенной формулы стремится к нулю при $n \rightarrow \infty$ и фиксированном m . В статистическом смысле $\mu = np$ представляет собой среднее значение числа появлений m события A при n испытаниях, так что $m - np$ есть отклонение числа появлений события A от его среднего значения. Выражение $\sigma = \sqrt{npq}$ пока можно воспринимать, как некоторый масштаб, в котором меряется это отклонение. Введя функцию

$$\varphi_0(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}$$

(также как и для тригонометрических функций для этой функции имеются специальные таблицы ее значений — эта функция табулирована), приближенную формулу Лапласа можно переписать в виде

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi_0(t) = \frac{1}{\sqrt{npq}} \varphi_0\left(\frac{m - np}{\sqrt{npq}}\right).$$

Пример. Вероятность поражения цели стрелком при одиночном выстреле равна $p = 0.2$. Какова вероятность того, что при 100 выстрелах цель будет поражена ровно 20 раз?

Здесь $p = 0.2$, $q = 0.8$, $n = 100$, $m = 20$. Отсюда

$$\sqrt{npq} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4, \quad t = \frac{m - np}{\sqrt{npq}} = \frac{20 - 100 \cdot 0.2}{4} = 0.$$

Кроме того, $\varphi_0(0) = 1/\sqrt{2\pi} \approx 0.40$. Поэтому, пользуясь формулой Лапласа,

$$P_{100}(20) \approx 0.40 \cdot \frac{1}{4} = 0.1.$$

Вычисления по точной формуле дают

$$P_{100}(20) = C_{100}^{20} p^{20} q^{80} \approx 0.0993002.$$

Интегральная теорема Лапласа.

Найдем вероятность $P_n(m_1, m_2)$ того, что в условиях схемы Бернулли событие A , имеющее вероятность $P(A) = p$, $0 < p < 1$, при n испытаниях появится не менее m_1 раз и не более m_2 раз? На основании теоремы сложения вероятности для несовместных событий получим

$$P_n(m_1, m_2) = \sum_{m=m_1}^{m_2} P_n(m).$$

По локальной теореме Лапласа приближенно

$$P_n(m_1, m_2) \approx \sum_{m=m_1}^{m_2} \frac{1}{\sqrt{npq}} \varphi_0(t_m), \quad \text{где } t_m = \frac{m - np}{\sqrt{npq}}, \quad \varphi_0(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

Так как

$$t_{m+1} - t_m = \frac{(m+1) - np}{\sqrt{npq}} - \frac{m - np}{\sqrt{npq}} = \frac{1}{\sqrt{npq}}, \quad \text{то}$$

$$P_n(m_1, m_2) \approx \sum_{m=m_1}^{m_2} \varphi_0(t_m) (t_{m+1} - t_m).$$

Полученная сумма является интегральной для функции $\varphi_0(t)$ на отрезке $[t_{m_1}, t_{m_2}]$. С увеличением n она приближенно соответствует определенному интегралу. Поэтому при достаточно больших n

$$P_n(m_1, m_2) \approx \int_{t_{m_1}}^{t_{m_2}} \varphi_0(t) dt = \frac{1}{\sqrt{2\pi}} \int_{t_{m_1}}^{t_{m_2}} e^{-t^2/2} dt, \quad \text{где } t_{m_1} = \frac{m_1 - np}{\sqrt{npq}}, \quad t_{m_2} = \frac{m_2 - np}{\sqrt{npq}}.$$

Это приближенное равенство и составляет содержание интегральной теоремы Лапласа. Введем стандартный интеграл вероятностей (функцию Лапласа)

$$\Phi_0(x) = \int_0^x \varphi_0(t) dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$

(первообразную для функции $\varphi_0(t)$). Тогда на основании формулы Ньютона-Лейбница получим еще одно представление формулы Лапласа

$$P_n(m_1, m_2) \approx \Phi_0(t_{m_2}) - \Phi_0(t_{m_1}).$$

Интеграл $\Phi_0(x)$ не выражается через элементарные функции, для его вычисления используются специальные таблицы, помещаемые в полных курсах по теории вероятностей. Кроме того, наряду с тригонометрическими и показательными функциями, эта функция может быть вычислена многими современными калькуляторами. Свойства функции $\Phi_0(x)$

1. $\Phi_0(0) = 0$,
2. $\Phi_0(+\infty) = 0.5$,
3. Эта функция нечетна, т.е. $\Phi_0(-x) = -\Phi_0(x)$,
4. $\Phi_0(x)$ — монотонно возрастающая функция (т.к. $\Phi_0'(x) = \varphi_0(x) > 0$). При $x > 3$, с точностью до тысячных, можно принять $\Phi_0(x) = 0.500$.

Пример. Вероятность поражения цели стрелком при одиночном выстреле равна $p = 0.2$. Какова вероятность того, что при 100 выстрелах цель будет поражена не менее 20 раз?

Здесь $p = 0.2$, $q = 0.8$, $20 \leq n \leq 100$. Отсюда

$$\sqrt{npq} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4, \quad t_{20} = \frac{20 - 100 \cdot 0.2}{4} = 0, \quad t_{100} = \frac{100 - 100 \cdot 0.2}{4} = 20.$$

Пользуясь интегральной формулой Лапласа, получим

$$P_{100}(20, 100) \approx \Phi_0(20) - \Phi_0(0) \approx 0.500.$$

Рассмотрим последовательность серий из n независимых испытаний с увеличивающимся n . Пусть каждая серия является схемой Бернулли с вероятностью $P(A) = p_n = \mu/n$ появления события A в каждом испытании из серии с общим числом испытаний n . Таким образом, в каждой серии испытаний среднее значение числа появлений событий A постоянно: $n \cdot p_n = \mu = \text{const}$. На основании биномиальной формулы для вероятности появления события A ровно m раз имеем выражение

$$P_n(m) = C_n^m \left(\frac{\mu}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^{n-m}.$$

Пусть m фиксировано и $n \rightarrow \infty$. Тогда

$$C_n^m \left(\frac{\mu}{n}\right)^m = \frac{n(n-1)\dots(n-m+1)}{m!n^m} \mu^m \rightarrow \frac{\mu^m}{m!}.$$

Кроме того,

$$\left(1 - \frac{\mu}{n}\right)^{n-m} = \left[\left(1 - \frac{\mu}{n}\right)^{n/\mu}\right]^\mu \left(1 - \frac{\mu}{n}\right)^{-m} \rightarrow e^{-\mu}.$$

Таким образом получаем

$$\lim_{n \rightarrow \infty} P_n(m) = \frac{\mu^m}{m!} e^{-\mu}.$$

Если n велико, то вероятность $P_n(m)$ сколь угодно мало отличается от полученного предела. Отсюда при больших n для искомой вероятности $P_n(m)$ имеем приближенную формулу Пуассона

$$P_n(m) \approx \frac{\mu^m}{m!} e^{-\mu},$$

где $\mu = np$ (теорема Пуассона). Вообще формулу Пуассона можно применять в случаях, когда число испытаний "велико", вероятность события $p_n = p$ мала, а $\mu = n \cdot p$ "не мало и не велико".

Пример. При выработке некоторой массовой продукции вероятность появления одного нестандартного изделия составляет 0.01. Какова вероятность, что в партии из 100 изделий этой продукции 2 изделия будут нестандартными?

Здесь вероятность $p = 0.01$ мала, число $n = 100$ велико, $\mu = n \cdot p = 100 \cdot 0.01 = 1$. Используя формулу Пуассона для искомой вероятности, получаем следующее значение

$$P_{100}(2) \approx \frac{\mu^2}{2!} e^{-\mu} = \frac{1}{2} e^{-1} \approx 0.184.$$

Лекция 4.

Случайные величины. Закон распределения и функция распределения дискретной случайной величины. Биномиальное распределение и распределение Пуассона.

Под случайной величиной понимают переменную, принимающую в результате испытания то или иное числовое значение в зависимости от случайного исхода испытания. Таким образом, случайная величина рассматривается как функция, аргументом которой служит элементарное случайное событие $w \in \Omega$ поля этого испытания. Среди случайных величин можно выделить два основных типа: величины дискретные и величины непрерывные.

Дискретной случайной величиной X (заданной на поле Ω пространства $(\Omega, \mathcal{A}, P)$) называется такая величина, которая может принимать конечное или бесконечное счетное множество числовых значений (т.е. такое множество, элементы которого могут быть занумерованы в каком-нибудь порядке и выписаны в последовательность $x_1, x_2, \dots, x_n, \dots$).

При дискретном распределении общая масса вероятности, равная 1, сосредоточена в счетной или конечной системе точек x_i . Другими словами, это – точечное распределение массы вероятности, подобное, например, точечному распределению электрических зарядов.

В качестве типичных примеров дискретных величин можно привести число дефектных изделий в выборке, число вызовов поступающих на телефонную станцию в тот или иной час работы, число остановок станка за одну смену и т.д.

С каждой дискретной случайной величиной X , принимающей значения x_n , связывают числа $p_n = P(X = x_n)$, обозначающие вероятность события, состоящего в том, что случайная величина X приняла значение x_n .

События $X = x_n$, рассмотренные при всех возможных значениях n , очевидно, образуют полную группу событий, поэтому

$$p_1 + p_2 + \dots + p_n + \dots = 1.$$

Соответствие между всеми возможными значениями дискретной величины и их вероятностями называется законом распределения данной случайной величины. В простейших случаях закон распределения случайной величины X удобно задавать таблицей, в которой первая строка содержит возможные значения случайной величины, а вторая – их вероятности.

Пример 1. В денежной лотерее разыгрывается 1 выигрыш в 1000 руб., 10 выигрышей по 100 руб., 100 выигрышей по 10 руб. при общем числе билетов 10000. Найти закон распределения случайного выигрыша X для владельца одного лотерейного билета.

Здесь возможные значения для X есть

$$x_1 = 1000, \quad x_2 = 100, \quad x_3 = 10, \quad x_4 = 0.$$

Им соответствуют вероятности

$$p_1 = 0.0001, \quad p_2 = 0.001, \quad p_3 = 0.01, \quad p_4 = 1 - (p_1 + p_2 + p_3) = 0.9889.$$

Пример 2. В серии повторных независимых испытаний, в каждом из которых данное событие A имеет одну и ту же вероятность $P(A) = p$ (схема Бернулли), число появлений m события A при n испытаниях можно рассматривать как случайную величину X со значениями $m = 0, 1, 2, \dots, n$. Закон распределения этой величины дается формулой Бернулли

$$p_m = P(X = m) = P_n(m) = C_n^m p^m q^{n-m},$$

где $p = P(A)$, $q = P(\bar{A}) = 1 - p$. Он называется биномиальным законом распределения случайной величины X . Про величину X при этом говорят, что она распределена по биномиальному закону с параметром n . По формуле бинома Ньютона легко проверить, что

$$\sum_{m=0}^n p_m = \sum_{m=0}^n C_n^m p^m q^{n-m} = (p + q)^n = 1.$$

В подавляющем числе задач теории вероятности в результате того или иного испытания $(\Omega, \mathcal{A}, P)$ наблюдается какая-либо числовая характеристика X этого испытания, т.е. случайная величина, связанная с этим испытанием. Обычно нет необходимости изучать вероятности всех возможных событий $A \in \mathcal{A}$, требуется лишь узнать вероятности событий, непосредственно связанных с наблюдаемой величиной X . Для дискретной случайной величины такие события и их вероятности задаются законом распределения. Если известен такой закон, то вероятность любого события, связанного со случайной величиной X , т.е. события, имеющего вид $X \in C$, $C \subset \mathbb{R}$, вычисляется по формуле

$$P(X \in C) = \sum_{n: x_n \in C} p(X = x_n) = \sum_{n: x_n \in C} p_n. \quad (1)$$

В частности, вероятность события $X \leq x$ определяется по формуле

$$F(x) = P(X \leq x) = \sum_{n: x_n \leq x} p_n.$$

Полученное соответствие $x \rightarrow F(x)$ принято называть функцией распределения случайной величины X . Нетрудно увидеть, что функция распределения обладает следующими характеристическими свойствами

- 1) $F(x)$ определена на всей числовой прямой и принимает значения на отрезке $[0, 1]$,
- 2) $F(-\infty) = 0$ – вероятность невозможного события равна нулю,
- 3) $F(x)$ неубывающая функция ($F(x_1) \leq F(x_2)$, если $x_1 < x_2$),
- 4) $F(+\infty) = 1$ – вероятность достоверного события равна 1.

Поскольку в той или иной вероятностной задаче основной интерес представляет именно ее наблюдение, т.е. некоторая числовая характеристика X , или, как мы ее назвали, случайная величина, то часто абстрагируются от конкретного вероятностного пространства $(\Omega, \mathcal{A}, P)$ и рассматривают только функцию распределения $F(x)$ этой случайной величины или ее закон распределения, т.е. последовательность

$p_n = P(X = x_n)$. При этом вероятности событий, связанных с X , подсчитываются по (1). Поэтому часто в той или иной задаче вместо указания пространства $(\Omega, \mathcal{A}, P)$ задают только закон распределения или функцию распределения случайной величины X и изучают связанные с ней события.

Пример 3. Говорят, что случайная величина X распределена по закону Пуассона (с параметром λ), если она является дискретной случайной величиной, принимающей целые неотрицательные значения $(0, 1, 2, \dots)$ с вероятностями

$$p_n = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, \dots$$

Согласно формуле Тейлора для функции e^x

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} e^{\lambda} = 1.$$

Лекция 5.

Функция распределения и плотность распределения непрерывной случайной величины, их взаимосвязь и свойства. Равномерное распределение вероятностей.

В самом общем случае числовая характеристика того или иного случайного испытания не является дискретной, т.е. все возможные ее наблюдения, вообще говоря, нельзя пересчитать и расположить в некоторый ряд чисел.

Простейшим примером такой характеристики является та, которая может принимать любое значение из некоторого отрезка на числовой прямой или вообще любое значение на числовой прямой.

Также как и для дискретной случайной величины, в общем случае числовую характеристику случайного испытания, заданную на вероятностном пространстве $(\Omega, \mathcal{A}, P)$, принято называть случайной величиной. Итак, случайные величины – это прежде всего функции $X : \Omega \rightarrow \mathbb{R}$.

Случайные величины рассматривают для изучения событий, связанных с этой величиной. Одним из самых простых событий, связанных с X , служит

$$(X < x) = \{\omega \in \Omega \mid X(\omega) < x\}, \quad x \in \mathbb{R},$$

– совокупность всех элементарных исходов из Ω , для которых $X(\omega)$ принимает значение меньшее чем x . По определению считают, что если X – некоторая случайная величина, то обязательно при каждом $x \in \mathbb{R}$ событие $(X < x)$ принадлежит $\mathcal{A}$. Вероятность этого события будем обозначать

$$F(x) = P(X < x),$$

а зависимость $F(x)$ от числа $x \in \mathbb{R}$ будем называть функцией распределения случайной величины X . Также как для дискретной случайной величины в общем случае функция распределения обладает следующими свойствами:

- 1) $F(x)$ определена на всей числовой прямой и принимает значения на отрезке $[0, 1]$,
- 2) $F(-\infty) = 0$ – вероятность невозможного события равна нулю,
- 3) $F(x)$ неубывающая функция ($F(x_1) \leq F(x_2)$, если $x_1 < x_2$). Действительно, если $x_1 < x_2$, то событие $(X < x_1)$ является подсобытием события $(X < x_2)$ и, значит,

$$P(X < x_1) \leq P(X < x_2)$$

по свойствам функции вероятности P ,

- 4) $F(+\infty) = 1$ – вероятность достоверного события равна 1.

Зная функцию распределения $F(x)$, можно для любого отрезка $[a, b]$ определить вероятность $P(a \leq X < b)$ события

$$(a \leq X < b) = \{\omega \in \Omega \mid X(\omega) \in [a, b)\}, \quad x \in \mathbb{R},$$

т.е. события, состоящего из всех элементарных исходов $\omega \in \Omega$, для которых числовое наблюдение $X(\omega)$ заключено в границах $[a, b)$. В самом деле, пусть A есть событие $X \in (-\infty, a)$, B есть событие $X \in (-\infty, b)$, C есть событие $X \in [a, b)$. Тогда, очевидно,

$$B = A \cup C, \quad A \cap C = \emptyset.$$

Поэтому по аксиоме сложения вероятностей $P(B) = P(A) + P(C)$, $P(C) = P(B) - P(A)$,

$$P(a \leq X < b) = F(b) - F(a),$$

причем $F(b) - F(a) \geq 0$ в силу свойства 3). Таким образом, вероятность того, что случайная величина X примет значение, принадлежащее промежутку $[a, b)$ равна приращению ее функции распределения на этом промежутке.

Случайная величина X называется непрерывной случайной величиной, если ее функция распределения непрерывна на оси $(-\infty, +\infty)$.

Теорема. Вероятность того, что непрерывная случайная величина примет заранее указанное значение a равна нулю.

Действительно, $P(X = a) \leq P(a \leq X < x) = F(x) - F(a)$ для $x > a$. В силу непрерывности функции F , имеем равенство

$$\lim_{x \rightarrow a} F(x) = F(a).$$

Поэтому

$$0 \leq P(X = a) \leq \lim_{x \rightarrow a} F(x) - F(a) = 0.$$

Следствие. Для непрерывной случайной величины справедливы равенства

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a), \quad a, b \in \mathbb{R}.$$

Действительно, по аксиоме сложения вероятностей,

$$P(a \leq X \leq b) = P(a \leq X < b) + P(X = b) = F(b) - F(a) + 0 = F(b) - F(a),$$

$$F(b) - F(a) = P(a \leq X < b) = P(a < X < b) + P(X = a) = P(a < X < b).$$

Частным случаем непрерывной случайной величины X является такая случайная величина, функция распределения которой имеет непрерывную производную

$$F'(x) = \varphi(x)$$

(или чуть общее – производную с конечным числом точек разрыва). Функцию $\varphi(x)$ называют плотностью вероятности (для данного распределения) или дифференциальным законом распределения случайной величины X .

Так как плотность вероятности является производной неубывающей функции $F(x)$, то она неотрицательна: $\varphi(x) \geq 0$. В то же время, в отличие от функции распределения, плотность вероятности может принимать значения большие единицы.

Так как $F(x)$ является первообразной для $\varphi(x)$, то по формуле Ньютона-Лейбница имеем

$$\int_a^b \varphi(x) dx = F(b) - F(a).$$

Поэтому

$$P(a \leq X \leq b) = \int_a^b \varphi(x) dx.$$

Геометрически эта вероятность представляет собой площадь криволинейной трапеции, ограниченной графиком плотности вероятности $y = \varphi(x)$, осью Ox и прямыми $x = a$, $x = b$. Полагая $a = -\infty$, $b = +\infty$, получаем достоверное событие $X \in (-\infty, +\infty)$, его вероятность равна 1:

$$1 = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} \varphi(x) dx. \quad (*)$$

Условие $(*)$ вместе с условием неотрицательности $\varphi(x) \geq 0$, $x \in \mathbb{R}$, являются характеристическими для того, чтобы функция $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ была бы плотностью вероятности некоторой случайной величины X .

Помимо возможности вычислить плотность вероятности $\varphi(x) = F'(x)$ по заданному закону распределения вероятностей $F(x)$, имеется и обратная возможность восстановить закон $F(c)$ по заданной плотности $\varphi(x)$:

$$F(c) = P(-\infty < X < c) = \int_{-\infty}^c \varphi(x) dx.$$

Простейшим примером непрерывной случайной величины служит равномерное распределение. Так называют случайную величину X , все возможные значения которой заполняют конечный промежуток $[a, b]$, плотность вероятности которой постоянна на $[a, b]$ и равна нулю вне $[a, b]$. Чтобы найти явный вид функции φ воспользуемся $(*)$:

$$1 = \int_{-\infty}^{\infty} \varphi(x) dx = \int_a^b \varphi(x) dx = c \int_a^b dx = c \cdot (b - a).$$

Таким образом, для равномерно распределенной на $[a, b]$ случайной величины X ее плотность вероятности $\varphi(x) = 1/(b - a)$, $x \in [a, b]$, $\varphi(x) = 0$, $x \notin [a, b]$.

Лекция 6.

Нормальный закон распределения вероятностей. Нормальная кривая. Функция Лапласа. Вычисление вероятности попадания в заданный интервал нормальной случайной величины. Правило трех сигм. Показательное распределение.

Распределение вероятностей случайной величины X называется нормальным, если ее плотность вероятности подчиняется закону Гаусса

$$\varphi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (1)$$

для любого значения $x \in \mathbb{R}$. Числа a и σ – произвольные числа (параметры распределения), причем σ положительно. График этой функции представляет из себя кривую в виде "горбика", принимающую наибольшее значение при $x = a$ и сосредоточенную, в основном, в окрестности этой точки (т.е. быстро убывающую к нулю при x отходящем от точки a). Распределение вероятностей случайной величины X имеет вид

$$F(x; a, \sigma) = \int_{-\infty}^x \varphi(t; a, \sigma) dt.$$

Легко видеть, что при любом x и любых значениях параметров распределения всегда $\varphi(x; a, \sigma) > 0$. С другой стороны полная площадь под всей кривой выражается интегралом

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(t-a)^2}{2\sigma^2}} dt,$$

который путем замены переменного t на $u = (t - a)/\sigma$ преобразуется в интеграл

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = 1. \quad (2)$$

Таким образом, площадь под нормальной кривой (при любых значениях параметров a и $\sigma > 0$) такова же, что и для нормальной кривой с параметрами $a = 0$ и $\sigma = 1$.

Из формулы (1) видно, что плотность распределения φ симметрична относительно ординаты $x = a$. Если изменять a , то кривая $y = \varphi(x; a, \sigma)$ будет перемещаться вдоль оси Ox , сохраняя свою форму. При $a = 0$ имеем семейство центрированных нормальных кривых

$$\varphi(x; 0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

зависящих от одного параметра σ .

При уменьшении параметра σ растет максимум $\frac{1}{\sqrt{2\pi}\sigma}$ функций $\varphi(x; a, \sigma)$. Этот подъем в центральной части компенсируется более резким спадом ее к оси Ox , так что общая величина площади под этой кривой остается неизменной (2). При очень малых значениях σ кривая становится похожей на тонкую иглу, направленную вдоль оси Oy . При этом почти вся площадь сконцентрирована на небольшом интервале с центром в a . При возрастании σ , наоборот, кривая принимает все более плосковершинную форму.

Чаще всего, однако, рассматривая случайную величину X , подчиненную нормальному закону с параметрами (a, σ) , переходят к нормированному распределению. Такой переход состоит в рассмотрении вместо величины X другой случайной величины

$$Z = \frac{X - a}{\sigma},$$

для которой

$$P(Z < z) = P\left(\frac{X - a}{\sigma} < z\right) = P(X < a + z\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{a+z\sigma} e^{-\frac{(u-a)^2}{2\sigma^2}} du.$$

Сделаем замену переменных в получившемся интеграле:

$$P(Z < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{v^2}{2}} dv.$$

Дифференцируя полученное равенство по переменной z , найдем плотность величины Z :

$$p_Z(z) = \varphi(z; 0, 1).$$

Тем самым, нормируя исходное распределение, получают стандартную случайную величину с параметрами $(0, 1)$. Все вопросы, связанные с нормальным распределением величины X , решают, переходя к вспомогательной величине Z , т.е. нормируя это распределение.

Для определения вероятности $P(x_1 < X < x_2)$ нахождения в интервале (x_1, x_2) случайной величины X , распределенной по нормальному закону с параметрами (a, σ) , требуется вычислять определенный интеграл

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{(t-a)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{(x_1-a)/\sigma}^{(x_2-a)/\sigma} e^{-\frac{v^2}{2}} dv.$$

Однако неопределенный интеграл вида

$$\int e^{-\frac{v^2}{2}} dv$$

не выражается через известные элементарные функции. В тоже время определенный интеграл

$$\int_{v_1}^{v_2} e^{-\frac{v^2}{2}} dv$$

при заданных v_1, v_2 может быть вычислен приближенно с любой степенью точности.

Определенный интеграл с переменным верхним пределом вида

$$\Phi_0(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{v^2}{2}} dv$$

носит название функции Лапласа. Свойства этой функции:

1. $\Phi_0(0) = 0$,
2. $\Phi_0(+\infty) = 1/2$, $\Phi_0(-\infty) = -1/2$,
3. $\Phi_0(-z) = -\Phi_0(z)$.

Для нахождения значений функции Лапласа пользуются специальными таблицами этой функции. С помощью функции Лапласа можно находить функцию распределения нормальной случайной величины с параметрами $(0, 1)$:

$$F_X(x; 0, 1) = \int_{-\infty}^0 \varphi(t; 0, 1) dt + \int_0^x \varphi(t; 0, 1) dt = 0.5 + \Phi_0(x),$$

а также вероятности произвольных событий, связанных со случайной величиной X , распределенной по нормальному закону с параметрами (a, σ) . Так, например,

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}} \int_0^{(x_2-a)/\sigma} e^{-\frac{v^2}{2}} dv - \frac{1}{\sqrt{2\pi}} \int_0^{(x_1-a)/\sigma} e^{-\frac{v^2}{2}} dv = \Phi_0\left(\frac{x_2-a}{\sigma}\right) - \Phi_0\left(\frac{x_1-a}{\sigma}\right).$$

Пользуясь полученным соотношением и таблицей для функции Φ_0 , можно легко определить вероятности попадания нормально распределенной с параметрами (a, σ) случайной величины X в интервалы $(a - \sigma, a + \sigma)$, $(a - 2\sigma, a + 2\sigma)$, $(a - 3\sigma, a + 3\sigma)$:

$$P(a - \sigma < X < a + \sigma) = \Phi_0(1) - \Phi_0(-1) = 2\Phi_0(1) \approx 0.68269 \approx 68\% \approx 2/3,$$

$$P(a - 2\sigma < X < a + 2\sigma) = 2\Phi_0(2) \approx 0.95450 \approx 95\%,$$

$$P(a - 3\sigma < X < a + 3\sigma) = 2\Phi_0(3) \approx 0.99730 \approx 99.7\%.$$

Рассматривая последний результат, видно, что вероятность нахождения величины X в интервале $(a - 3\sigma, a + 3\sigma)$ близка к 1. Поэтому "трехсигмовые" границы $a \pm 3\sigma$ принимаются за границы практически предельных возможных значений величины X (правило трех сигм).

Пример. Найти вероятность того, что отклонение размера валика от номинала находится в пределах $(-0.11\text{мм}, -0.07\text{мм})$, если предполагается, что это отклонение распределено по нормальному закону с параметрами $(a = -0.112\text{мм}, \sigma = 0.043\text{мм})$:

$$P(-0.11 < X < -0.07) = \Phi_0\left(\frac{0.042}{0.043}\right) - \Phi_0\left(\frac{0.002}{0.043}\right) = \Phi_0(0.98) - \Phi_0(0.05) \approx 0.3166.$$

Еще одним примером непрерывной случайной величины, задаваемой при помощи функции плотности, служит случайная величина X , распределенная по показательному закону с параметром $\lambda > 0$. Плотность вероятности X задается формулой $p(x) = \lambda e^{-\lambda x}$ при $x \geq 0$, $p(x) = 0$ при $x < 0$. Тем самым выполняется условие $p(x) \geq 0$. Проверим условие нормировки:

$$\int_{\mathbf{R}} p(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \int_0^{\infty} d(-e^{-\lambda x}) = -e^{-\infty} - (-e^0) = 1.$$

С помощью показательного распределения обычно задается распределение случайных отрезков времени между последовательными наступлениями редких событий. Функция распределения этой случайной величины имеет вид

$$F(x) = P(X < x) = \int_{-\infty}^x p(x) dx$$

и, значит, $F(x) = 0$, если $x \leq 0$, $F(x) = 1 - e^{-\lambda x}$, если $x > 0$.

Вообще выбор того или иного закона распределения вероятности из общего числа модельных законов (запас которых достаточно велик) производится по предварительным статистическим наблюдениям той или иной случайной величины и по анализу полученных результатов с целью выбрать наиболее подходящий закон.

Лекция 7.

Основные числовые характеристики дискретных и непрерывных случайных величин: математическое ожидание, дисперсия, среднее квадратическое отклонение, моменты. Их свойства и примеры.

Основные числовые характеристики случайных величин будем определять отдельно для дискретных и непрерывных случайных величин. Пусть X – дискретная случайная величина, $\{x_i\}$ – полный перечень ее возможных значений, $\{p_i\}$ – закон распределения вероятностей этой случайной величины. Математическим ожиданием X называют

$$M(X) = \sum_i p_i x_i \quad (1)$$

(если величина X принимает бесконечное число значений, то в этом определении предполагается сходимость ряда $\sum p_i x_i$). Математическое ожидание $M(X)$, таким образом, есть величина постоянная и поэтому представляет числовую характеристику случайной величины X , некоторое ее среднее значение: если $\underline{x} \leq x_i \leq \bar{x}$ при всех возможных i , то

$$\underline{x} = \sum_i p_i \underline{x} \leq \sum_i p_i x_i \leq \sum_i p_i \bar{x} = \bar{x}$$

и, значит, $\underline{x} \leq M(X) \leq \bar{x}$.

Пример. X — число появлений события A в данном испытании, $P(A) = p$, $P(\bar{A}) = 1 - p = q$. Тогда случайная величина X принимает два значения: 1 с вероятностью p и 0 с вероятностью q . По определению

$$M(X) = 1 \cdot p + 0 \cdot q = p.$$

Для непрерывной случайной величины X с плотностью вероятности $p(x)$ математическое ожидание определяется подобной (1) формулой

$$M(X) = \int_{\mathbf{R}} x p(x) dx. \quad (2)$$

Пример. X — равномерно распределенная на отрезке $[a, b]$ случайная величина, $p(x) = 1/(b - a)$, $x \in [a, b]$, $p(x) = 0$, $x \notin [a, b]$, — ее плотность вероятности. Тогда

$$M(X) = \int_a^b x p(x) dx = \frac{b^2 - a^2}{2(b - a)} = \frac{b + a}{2}.$$

Основные свойства математического ожидания для дискретной и непрерывной случайных величин выглядят одинаково. Однако для упрощения обоснования этих формул будем производить лишь для дискретных случайных величин.

Пусть X, Y — пара дискретных случайных величин, $\{x_i\}, \{y_j\}$, $i \in I, j \in J$, — их возможные значения. Суммой $X + Y$ этих величин называется дискретная случайная величина, принимающая значения $\{x_i + y_j\}$, $i \in I, j \in J$ для всех возможных $i \in I, j \in J$. При этом событие $(X + Y = x_i + y_j)$, с выделенными индексами i, j , вообще говоря подразделяется на несколько подсобытий, среди которых обязательно присутствует событие $(X = x_i, Y = y_j)$ с вероятностью наступления

$$P(X = x_i) \cdot P(Y = y_j \mid X = x_i).$$

С другой стороны помимо этого, числовое значение $x_i + y_j$ может оказаться равным значению $x_m + y_n$ для других значений индексов $m \neq i, n \neq j$. Это означает, что событие $(X + Y = x_i + y_j)$ в качестве подсобытий содержит все возможные $(X = x_m, Y = y_n)$, для которых $x_m + y_n = x_i + y_j$, $m \in I, n \in J$. Таким образом, событие $(X + Y = x_i + y_j)$ устроено достаточно сложно, вероятность его наступления может быть записана в виде

$$p_{ij} = \sum_{x_m + y_n = x_i + y_j} P(X = x_m) \cdot P(Y = y_n \mid X = x_m). \quad (3)$$

Точно также разностью, произведением или частным случайных величин X и Y называется случайная величина, принимающая значения $\{x_i - y_j\}, \{x_i \cdot y_j\}$ или $\{x_i / y_j\}$. Вероятность наступления соответствующих событий подсчитывается по формулам, аналогичным (3) (в последнем случае предполагается, что случайная величина Y не принимает нулевого значения).

Еще одним важным понятием, позволяющим во многих случаях существенно упростить рассмотрения, является понятие независимости случайных величин. Величины X и Y называются независимыми, если возможные значения и закон распределения каждой из них один и тот же при любом выборе допустимых значений другой:

$$P(Y = y_j \mid X = x_i) = P(Y = y_j), \quad P(X = x_i \mid Y = y_j) = P(X = x_i).$$

В противном случае они называются зависимыми.

Пара непрерывных случайных величин X и Y , с функциями распределения $F(x)$ и $G(y)$, называются независимыми, если для их совместной функции распределения выполняются равенства:

$$H(x, y) = P(X < x, Y < y) = P(X < x)P(Y < y) = F(x)G(y).$$

По аналогии несколько случайных величин называются взаимно независимыми, если возможные значения и законы распределения любой из них не зависят от того, какие возможные значения приняли остальные случайные величины.

Теорема 1. Математическое ожидание постоянной величины равно этой постоянной $M(C) = C$.

Доказательство. Постоянная величина является дискретной случайной величиной, принимающее единственное значение C с вероятностью 1. Поэтому по определению $M(C) = C \cdot 1 = C$.

Теорема 2. Математическое ожидание суммы двух (или большего числа) случайных величин X, Y равно сумме математических ожиданий этих величин:

$$M(X + Y) = M(X) + M(Y).$$

Доказательство. Пусть $\{x_i\}, \{y_j\}, i \in I, j \in J$, – все возможные значения пары случайных величин X и Y . Нетрудно видеть, что математическое ожидание суммы случайных величин может быть записано в виде

$$\begin{aligned} M(X + Y) &= \sum_{i \in I} \sum_{j \in J} (x_i + y_j) P(X = x_i) P(Y = y_j | X = x_i) = \\ &= \sum_{i \in I} \sum_{j \in J} x_i P(X = x_i) P(Y = y_j | X = x_i) + \sum_{i \in I} \sum_{j \in J} y_j P(X = x_i) P(Y = y_j | X = x_i) = \\ &= \sum_{i \in I} x_i P(X = x_i) \sum_{j \in J} P(Y = y_j | X = x_i) + \sum_{j \in J} y_j \sum_{i \in I} P(Y = y_j) P(X = x_i | Y = y_j) = \\ &= \sum_{i \in I} x_i P(X = x_i) + \sum_{j \in J} y_j P(Y = y_j) = M(X) + M(Y), \end{aligned}$$

т.к., например,

$$\sum_{j \in J} P(Y = y_j | X = x_i) = P(\cup_{j \in J} (Y = y_j) | X = x_i) = P(\Omega | X = x_i) = 1.$$

Для нескольких случайных величин, например для трех X, Y и Z , имеем

$$M(X + Y + Z) = M(X + Y) + M(Z) = M(X) + M(Y) + M(Z).$$

Следствие. Если C – постоянная величина, то $M(X + C) = M(X) + C$.

Теорема 3. Математическое ожидание произведения двух независимых случайных величин X и Y равно произведению математических ожиданий этих величин:

$$M(X \cdot Y) = M(X) \cdot M(Y).$$

Пусть $\{x_i, p_i\}, i \in I, \{y_j, q_j\}, j \in J$, – законы распределения соответственно случайных величин X и Y . Тогда, по аналогии с теоремой 2,

$$\begin{aligned} M(X \cdot Y) &= \sum_{i \in I} \sum_{j \in J} (x_i \cdot y_j) P(X = x_i) P(Y = y_j | X = x_i) = \sum_{i \in I} \sum_{j \in J} (x_i \cdot y_j) P(X = x_i) P(Y = y_j) = \\ &= \sum_{i \in I} x_i p_i \sum_{j \in J} y_j q_j = M(X) \cdot M(Y). \end{aligned}$$

Следствие. Математическое ожидание произведения нескольких взаимно независимых случайных величин равно произведению математических ожиданий этих величин.

Действительно, например для трех взаимно независимых случайных величин X, Y и Z , имеем

$$M(X \cdot Y \cdot Z) = M(X \cdot Y) \cdot M(Z) = M(X) \cdot M(Y) \cdot M(Z).$$

Следствие. Постоянный множитель можно выносить за знак математического ожидания.

Если C – постоянная величина, X – произвольная случайная величина, то C и X независимы. Поэтому

$$M(C \cdot X) = M(C) \cdot M(X) = C \cdot M(X).$$

Следствие. Математическое ожидание разности двух случайных величин X, Y равно разности математических ожиданий этих величин: $M(X - Y) = M(X) - M(Y)$.

Действительно,

$$M(X - Y) = M(X) + M(-Y) = M(X) + (-1)M(Y) = M(X) - M(Y).$$

Пусть X – случайная величина, $M(X)$ – ее математическое ожидание (среднее значение). Случайную величину $X - M(X)$ называют отклонением.

Теорема 4. Для любой случайной величины X математическое ожидание ее отклонения равно нулю:

$$M(X - M(X)) = 0.$$

Доказательство. Так как $M(X)$ – постоянна, то

$$M(X - M(X)) = M(X) - M(M(X)) = M(X) - M(X) = 0.$$

Определение. Дисперсией (рассеянием) случайной величины X называют

$$D(X) = M(\{X - M(X)\}^2).$$

Из определения следует, что дисперсия случайной величины постоянна, т.е. является числовой характеристикой этой величины. Если случайная величина имеет закон распределения вероятности $\{x_i, p_i\}$, $i \in I$, то, обозначая $\mu = M(X)$, получим также

$$D(X) = \sum_{i \in I} (x_i - \mu)^2 p_i.$$

Корень квадратный из дисперсии называется средним квадратичным отклонением

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

Пример. Пусть $(4, 1/4)$, $(10, 1/2)$, $(20, 1/4)$ – закон распределения дискретной случайной величины X . Тогда

$$M(X) = 4 \cdot 1/4 + 10 \cdot 1/2 + 20 \cdot 1/4 = 11,$$

$$D(X) = (4 - 11)^2 \cdot 1/4 + (10 - 11)^2 \cdot 1/2 + (20 - 11)^2 \cdot 1/4 = 33, \quad \sigma(X) = \sqrt{D(X)} \approx 5,75.$$

Дисперсия служит мерой рассеяния (разброса) значений случайной величины X : для малой дисперсии характерно, что значения этой случайной величины концентрируются около ее математического ожидания, а далеко от математического ожидания могут отстоять лишь значения, вероятность появления которых мала. Если же дисперсия велика, то концентрация вокруг какого-либо центра исключается.

Теорема 5. X – случайная величина. Тогда

$$D(X) = M(X^2) - (M(X))^2.$$

Доказательство.

$$\begin{aligned} D(X) &= M(\{X - M(X)\}^2) = M(X^2 - 2X \cdot M(X) + (M(X))^2) = \\ &= M(X^2) - M(2X \cdot M(X)) + M((M(X))^2) = M(X^2) - (M(X))^2. \end{aligned}$$

Теорема 6. Дисперсия постоянной величины равна нулю.

Доказательство. Пусть C – постоянная величина. Тогда

$$D(C) = M(\{C - M(C)\}^2) = M(\{C - C\}^2) = M(0) = 0.$$

Теорема 7. Дисперсия суммы независимых случайных величин X, Y равна сумме дисперсий этих величин:

$$D(X + Y) = D(X) + D(Y).$$

Действительно,

$$\begin{aligned} D(X + Y) &= M(\{X + Y - M(X + Y)\}^2) = M(\{(X - M(X)) + (Y - M(Y))\}^2) = \\ &= M(\{X - M(X)\}^2) + 2M((X - M(X))(Y - M(Y))) + M(\{Y - M(Y)\}^2). \end{aligned}$$

Так как X, Y – независимые случайные величины, то $X - M(X), Y - M(Y)$ – также независимые случайные величины. Поэтому

$$M((X - M(X))(Y - M(Y))) = M(X - M(X))M(Y - M(Y)) = 0.$$

Значит, $D(X + Y) = D(X) + D(Y)$.

Следствие. Дисперсия суммы нескольких независимых случайных величин равна сумме дисперсий этих величин.

Следствие. Если C – постоянная величина, X – произвольная случайная величина, то

$$D(X + C) = D(X).$$

Следствие. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(C \cdot X) = C^2 D(X).$$

Действительно,

$$D(C \cdot X) = M(C^2 X^2) - (M(CX))^2 = C^2 M(X^2) - (CM(X))^2 = C^2 D(X).$$

Следствие. Дисперсия разности двух независимых случайных величин X, Y равна сумме дисперсий этих величин: $D(X - Y) = D(X) + D(Y)$.

Доказательство.

$$D(X - Y) = D(X) + D(-Y) = D(X) + (-1)^2 D(Y) = D(X) + D(Y).$$

Пример. В схеме испытаний Бернулли определить математическое ожидание и дисперсию для числа X появления события A при n независимых испытаниях, в каждом из которых вероятность $P(A) = p$ появления события A постоянна.

Случайная величина X принимает значения $0, 1, \dots, n$ и распределена по биномиальному закону:

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

Используя этот закон распределения, математическое ожидание и дисперсию X можно вычислить непосредственно по определению. Применим, однако, установленные теоремы. Случайную величину X можно рассматривать как сумму независимых случайных величин

$$X = X_1 + \dots + X_n,$$

где X_k – число появлений события A в k -ом испытании. Эта величина принимает только 2 значения: 0 с вероятностью q и 1 с вероятностью p . Поэтому

$$M(X_k) = 0 \cdot q + 1 \cdot p = p, \quad D(X_k) = (0 - p)^2 \cdot q + (1 - p)^2 \cdot p = p^2 q + q^2 p = pq.$$

Значит, по теореме о математическом ожидании суммы случайных величин

$$M(X) = M(X_1) + \dots + M(X_n) = n \cdot p,$$

по теореме о дисперсии суммы независимых случайных величин

$$D(X) = D(X_1) + \cdots + D(X_n) = n \cdot pq.$$

Поэтому среднее квадратичное уклонение (стандарт)

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

В заключение рассмотрим понятие моментов для случайной величины $X : \Omega \rightarrow \mathbb{R}$. Предположим $f : \mathbb{R} \rightarrow \mathbb{R}$ — некоторая непрерывная функция. Рассмотрим случайную величину $f(X) : \Omega \rightarrow \mathbb{R}$, определенную как композиция двух функций

$$f(X)(w) = f(X(w)), \quad w \in \Omega.$$

Для этой случайной величины также можно рассмотреть математическое ожидание $M(f(X))$. При специальном выборе функции f это математическое ожидание носит особое название. Пусть $f(x) = x^k$, $k \in \mathbb{Z}_+$. Величина $M(X^k)$ называется моментом порядка k , величина $M((X - M(X))^k)$ называется центральным моментом порядка k .

Примеры. Пусть X — случайная величина, распределенная по показательному закону с параметром λ : $p(x) = \lambda e^{-\lambda x}$ при $x \geq 0$, $p(x) = 0$ при $x < 0$. Вычислим ее математическое ожидание и дисперсию.

$$M(X) = \int_{\mathbb{R}} x \cdot p(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = -x \cdot e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = 1/\lambda$$

и, аналогично,

$$M(X^2) = \int_{\mathbb{R}} x^2 \cdot p(x) dx = 2/\lambda^2.$$

Поэтому $D(X) = M(X^2) - (M(X))^2 = 1/\lambda^2$.

Пусть X — нормальная случайная величина с параметрами (a, σ) : $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$, $x \in \mathbb{R}$. Тогда

$$M(X) = \int_{\mathbb{R}} x \cdot p(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (x - a + a) \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} a \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = a.$$

$$D(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (x - a)^2 \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^2 \cdot e^{-\frac{t^2}{2}} dt = \sigma^2.$$

Таким образом параметры (a, σ^2) для нормально распределенной случайной величины являются ее математическим ожиданием и дисперсией.

Лекция 8.

Случайные векторы. Закон распределения вероятностей дискретной двумерной случайной величины. Функция распределения и плотность распределения двумерной случайной величины, их свойства. Вероятность попадания случайной точки в произвольную область. Отыскание плотностей вероятности составляющих двумерной случайной величины. Условное распределение вероятности, регрессия.

Во многих практических задачах наблюдение того или иного испытания представляет собой не одномерное число, а некоторое конечное их количество: два числа, три числа или более. В этом случае вместо термина случайная величина используют название случайный вектор той или иной размерности. Как и для случайных величин ограничимся рассмотрением дискретных случайных векторов и непрерывных случайных векторов, обладающих плотностью вероятности.

Дискретные случайные вектора принимают конечное или счетное число значений. Для определенности будем рассматривать двумерные случайные вектора. В этом случае каждое значение случайного вектора представляет из себя пару чисел (x, y) . При этом каждую координату можно трактовать как реализацию некоторой дискретной случайной величины: X — для первой координаты, Y — для второй. Также как и для

случайных величин, с каждым случайным вектором связывают закон распределения, называемый совместным законом распределения вероятностей величин X и Y . Этот закон представляет из себя соответствие, которое каждой паре (x_i, y_j) значений случайного вектора ставит в соответствие неотрицательное число

$$p_{i,j} = P(X = x_i, Y = y_j).$$

При этом предполагается, что все комбинации $(X = x_i, Y = y_j)$ составляют полную группу событий и потому сумма вероятностей

$$\sum_i \sum_j p_{i,j} = 1.$$

Если просуммировать все вероятности с постоянным значением i , то по правилу сложения несовместных событий

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

получим вероятность события $(X = x_i)$. Таким же образом

$$\sum_i P(X = x_i, Y = y_j) = P(\cup_i (X = x_i), Y = y_j) = P(Y = y_j) = p(y_j).$$

Тем самым одномерные законы или таблицы распределения каждой величины X и Y в отдельности полностью определяются, если известна таблица распределения двумерной величины.

Раньше уже отмечалось, что зависимость между двумя случайными событиями выражается в том, что условная вероятность одного события при наступлении другого события отличается от безусловной вероятности первого. Аналогично этому, чтобы исследовать влияние одной величины на изменение другой величины, рассматривают условные законы распределения первой величины при фиксированных значениях второй: если величина X получила одно из своих значений x_i , то другая величина, Y , может принять любое из своих возможных значений $y_1, y_2, \dots$; однако вероятности этих значений будут, вообще говоря, отличаться от вероятностей $p(y_1), p(y_2), \dots$. Эти вероятности мы называли условными и обозначали

$$p(y_j | x_i) = P(Y = y_j | X = x_i) = p(x_i, y_j) / p(x_i).$$

Совокупность условных вероятностей $p(y_1 | x_i), p(y_2 | x_i), \dots$, отвечающих одному и тому же условию $X = x_i$ называют условным распределением Y при $X = x_i$. Очевидно,

$$\sum_j p(y_j | x_i) = \sum_j p(y_j, x_i) / p(x_i) = 1.$$

Для описания условных законов распределения вероятности используются те же характеристики, что и для безусловных случайных величин. Наиболее важной характеристикой является условное математическое ожидание $M(Y | x)$ величины Y при фиксированном значении $X = x$, где x может равняться $x_1, x_2, \dots$. Оно определяется равенством

$$\bar{y}(x) = M(Y | x) = \sum_j y_j p(y_j | x).$$

Аналогично вводятся условная дисперсия и условные моменты более высоких порядков.

Зависимость $\bar{y}(x) = M(Y | x)$ среднего значения случайной величины Y от значения, которое приняла случайная величина $X = x$, называется регрессией Y по X .

Аналогичным образом вводятся $p(x_i | y_j)$ – условные законы величины X при фиксированных значениях $Y = y_j$, условные числовые характеристики и функция регрессии $\bar{x}(y) = M(X | y)$ случайной величины X по Y .

В случае непрерывного распределения величин X и Y их совместное распределение задается с помощью плотности вероятности $p(x, y)$ – неотрицательной интегрируемой функции двух переменных (x, y) на плоскости:

$$\int_{\mathbb{R}^2} p(x, y) dx dy = 1.$$

При этом наличие функции плотности вероятности по определению означает, что вероятность случайной точке Q со случайными координатами (X, Y) попасть в какую-либо область $C \subset \mathbb{R}^2$ выражается равенством

$$P(Q(X, Y) \in C) = \int_C p(x, y) dx dy.$$

Функция

$$P_{XY}(x, y) = P(X < x, Y < y) = \int_{-\infty}^x \int_{-\infty}^y p(x, y) dx dy$$

называется интегральной функцией распределения двумерной величины (X, Y) .

Свойства интегральной функции: эта функция возрастает при возрастании каждого из переменных x и y и стремится к единице, когда оба аргумента неограниченно возрастают.

Наиболее простым примером является равномерное распределение двумерной величины на на какой-нибудь части плоскости C . В этом случае $p(x, y) = c > 0$ для $(x, y) \in C$ и $p(x, y) = 0$ для $(x, y) \notin C$. Константа c определяется из условия

$$\int_{\mathbb{R}^2} p(x, y) dx dy = c \int_C dx dy = c \cdot S_C = 1, \quad c = 1/S_C,$$

где S_C – площадь области C на плоскости. Вероятность попасть на какую-нибудь площадку G равна

$$\int_G p(x, y) dx dy = 1/S_C \int_{C \cap G} dx dy = \frac{S_{C \cap G}}{S_C}.$$

Если плотность $p(x, y)$ двумерного распределения задана, то можно определить одномерные функции распределения и плотности. Действительно,

$$P(X < x) = P(X < x, Y < \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} p(x, y) dx dy$$

и, значит, для плотности $p_X(x)$ величины X имеем

$$p_X(x) = \int_{-\infty}^{\infty} p(x, y) dy.$$

Эта формула вполне аналогична формулам, полученным для дискретных случайных величин. Точно также

$$p_Y(y) = \int_{-\infty}^{\infty} p(x, y) dx.$$

Условные законы распределения Y при заданном значении $X = x$ определяются в данном случае условной плотностью

$$p_Y(y | x) = \frac{p(x, y)}{p_X(x)} = \frac{p(x, y)}{\int_{\mathbb{R}} p(x, y) dy}$$

и, аналогично, для условной плотности X при заданном $Y = y$

$$p_X(x | y) = \frac{p(x, y)}{p_Y(y)} = \frac{p(x, y)}{\int_{\mathbb{R}} p(x, y) dx}.$$

Линии регрессии Y по X и X по Y определяются теперь как условные математические ожидания

$$\bar{y}(x) = M(Y | x) = \int_{\mathbb{R}} y \cdot p(y | x) dy = \frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy},$$

$$\bar{x}(y) = M(X | y) = \int_{\mathbb{R}} x \cdot p(x | y) dx = \frac{\int_{\mathbb{R}} x \cdot p(x, y) dx}{\int_{\mathbb{R}} p(x, y) dx}.$$

Все эти понятия могут быть легко обобщены на случай любого числа измерений.

Рассмотренные связи между случайными величинами позволяют ввести важное понятие о независимых величинах. Для дискретных случайных величин определение независимости состоит в выполнении равенств

$$p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$$

при всех возможных исходах (x, y) испытания (X, Y) . Для таких случайных величин

$$p(y | x) = p_Y(y)$$

– не зависит от значения x , принятого случайной величиной X . Поэтому, в частности, все условные математические ожидания $M(Y | x)$ величины Y также не будут зависеть от x и будут равны безусловному математическому ожиданию Y :

$$\bar{y}(x) = M(Y).$$

Аналогично $\bar{x}(y) = M(X)$.

Для непрерывных случайных величин X и Y их независимость по определению означает выполнение равенств

$$P(X < x, Y < y) = P(X < x) \cdot P(Y < y),$$

выполняющихся при всех возможных значениях $(x, y) \in \mathbb{R}^2$. Это определение можно переписать в терминах функций распределения

$$P_{XY}(x, y) = P_X(x) \cdot P_Y(y)$$

и в терминах плотностей распределения

$$p_{XY}(x, y) = p_X(x) \cdot p_Y(y).$$

Лекция 9.

Числовые характеристики двумерных случайных величин: начальные и центральные моменты. Корреляционный момент и коэффициент корреляции. Коррелированность и зависимость случайных величин. Нормальный закон распределения на плоскости.

Рассмотренные связи между случайными величинами позволяют ввести важное понятие о независимых величинах. Для дискретных случайных величин определение независимости состоит в выполнении равенств

$$p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$$

при всех возможных исходах (x, y) испытания (X, Y) . Для таких случайных величин

$$p(y | x) = p_Y(y)$$

– не зависит от значения x , принятого случайной величиной X . Поэтому, в частности, все условные математические ожидания $M(Y | x)$ величины Y также не будут зависеть от x и будут равны безусловному математическому ожиданию Y :

$$\bar{y}(x) = M(Y).$$

Аналогично $\bar{x}(y) = M(X)$.

Для непрерывных случайных величин X и Y их независимость по определению означает выполнение равенств

$$P(X < x, Y < y) = P(X < x) \cdot P(Y < y),$$

выполняющихся при всех возможных значениях $(x, y) \in \mathbb{R}^2$. Это определение можно переписать в терминах функций распределения

$$P_{XY}(x, y) = P_X(x) \cdot P_Y(y)$$

и в терминах плотностей распределения

$$p_{XY}(x, y) = p_X(x) \cdot p_Y(y).$$

Пусть X и Y – пара случайных величин, $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ функция двух переменных. Математическое ожидание $M(\varphi(X, Y))$ случайной величины $\varphi(X, Y)$ определяется как

$$M(\varphi(X, Y)) = \sum_i \sum_j \varphi(x_i, y_j) p(x_i, y_j)$$

для дискретных случайных величин и как

$$M(\varphi(X, Y)) = \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} \varphi(x, y) p(x, y) dx dy$$

для непрерывных случайных величин. В частности, если $\varphi(x, y) = x + y$, то $M(\varphi(X, Y)) = M(X) + M(Y)$. Еще раз рассмотрим вопрос о дисперсии суммы двух случайных величин. Если эти величины независимы, то, как было показано,

$$D(X + Y) = D(X) + D(Y).$$

Если же эти величины зависимы, то такое равенство вообще говоря не выполняется, выражение

$$\text{cov}(X, Y) = (D(X + Y) - D(X) - D(Y))/2$$

называется ковариацией пары случайных величин и служит некоторой характеристикой зависимости X и Y . Для этой характеристики справедливо равенство

$$\begin{aligned} \text{cov}(X, Y) &= M((X - M(X)) \cdot (Y - M(Y))) = \\ &= M(XY) - M(X \cdot M(Y)) - M(Y \cdot M(X)) + M(X) \cdot M(Y) = M(XY) - M(X) \cdot M(Y). \end{aligned}$$

Величина $\text{cov}(X, Y)$ зависит от единиц измерения, в которых выражают X и Y , поэтому она сама по себе еще не может служить показателем их связи. Чтобы иметь дело с безразмерным показателем, рассматривают ковариации нормированных отклонений

$$X^* = \frac{X - M(X)}{\sigma(X)}, \quad Y^* = \frac{Y - M(Y)}{\sigma(Y)}.$$

Каждая из них имеет центром нуль и дисперсию равную единице,

$$\text{cov}(X, Y) = \text{cov}(X^*, Y^*) = M(X^*, Y^*) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Число $\text{cor}(X, Y)$ называется коэффициентом корреляции величин X и Y . Для независимых случайных величин коэффициент корреляции равен нулю, так как для них $\text{cov}(X, Y) = 0$. Однако обратного заключения сделать нельзя: величины могут быть связаны функционально, а коэффициент их корреляции при этом будет равен нулю: пусть, например, случайная величина X симметрично распределена относительно начала координат (т.е. для плотности вероятности $p(x) = p(-x)$). Тогда $M(X) = 0$ и для случайной величины $Y = X^2$, связанной с X функциональной зависимостью,

$$\text{cov}(X, Y) = M(X \cdot Y) - M(X) \cdot M(Y) = M(X^3) - 0 = 0.$$

Рассмотрим теперь другой крайний случай. Вычислим дисперсию суммы и разности величин X^* и Y^* . В этом случае $D(X^*) = D(Y^*) = 1$ и $\text{cov}(X^*, Y^*) = \text{cor}(X, Y)$. Поэтому

$$0 \leq D(X^* \pm Y^*) = 1 + 1 \pm 2\text{cor}(X, Y) = 2(1 \pm \text{cor}(X, Y)).$$

Полученное означает, что всегда

$$-1 \leq \text{cor}(X, Y) \leq 1.$$

Кроме того, если $\text{cor}(X, Y) = 1$, то $D(X^* + Y^*) = 0$, что означает, что случайная величина $X^* + Y^*$ принимает одно значение с вероятностью 1, т.е. случайные величины X и Y связаны линейной зависимостью. Точно также эти величины связаны линейной зависимостью и в случае $\text{cor}(X, Y) = -1$.

Для суммы большего числа случайных величин число коэффициентов ковариации и корреляции будет расти: для каждой пары – своя пара коэффициентов. Чтобы ввести порядок, их размещают в таблицу: каждый коэффициент – на место в матрице согласно индексам двух случайных величин, образующих этот коэффициент. Сами матрицы называют матрицами корреляции и ковариации.

Следующее наблюдение оказывается чрезвычайно важным в случае наблюдения последовательности независимых испытаний (оно оказывается верным и для попарно некоррелированных случайных величин): пусть $X_1, \dots, X_n$ – последовательность независимых испытаний. Тогда

$$D(X_1 + \dots + X_n) = D(X_1) + \dots + D(X_n), \quad \sigma(X_1 + \dots + X_n) = \sqrt{\sigma(X_1)^2 + \dots + \sigma(X_n)^2}.$$

Для случая когда все величины X_i имеют одинаковую дисперсию σ^2 , эти равенства переписутся как

$$D(X_1 + \dots + X_n) = n\sigma^2, \quad \sigma(X_1 + \dots + X_n) = \sqrt{n}\sigma.$$

Это означает, что рост дисперсии суммы происходит пропорционально числу слагаемых, а среднее квадратичное отклонение растет пропорционально квадратному корню из числа слагаемых. Важным следствием этого являются формулы для дисперсии средней арифметической

$$D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}D(X_1 + \dots + X_n) = \frac{\sigma^2}{n}, \quad \sigma\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}.$$

Итак, если все величины X_i имеют одинаковую дисперсию σ^2 (например, при рассмотрении n повторных независимых измерений одной и той же величины), то среднее квадратичное отклонение средней арифметической величины для n величин будет в $\sqrt{n}$ раз меньше, чем для одной величины. Для измерений одной величины так сказывается компенсация в сторону плюс и в сторону минус от фактического измеряемого значения, которая имеет место при сложении результатов отдельных измерений.

Важным случаем двумерного распределения для практических приложений служит двумерное нормальное распределение. В этом случае плотность распределения задается выражением

$$p(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x, y)}, \quad (1)$$

где

$$Q(x, y) = \frac{1}{1-\rho^2} \left(\frac{(x - M(X))^2}{\sigma_X^2} + \frac{(y - M(Y))^2}{\sigma_Y^2} - 2\rho \frac{(x - M(X))(y - M(Y))}{\sigma_X\sigma_Y} \right),$$

σ_X, σ_Y – средние квадратичные отклонения случайных величин X и Y (т.е. корни из их дисперсий), $M(X)$ и $M(Y)$ – математические ожидания этих величин, $\rho = \text{cor}(X, Y)$ – их коэффициент корреляции.

Если случайные величины X и Y независимы и нормально распределены с параметрами $(M(X), \sigma_X^2)$, $(M(Y), \sigma_Y^2)$, то их коэффициент корреляции $\rho = 0$; в этом случае нетрудно видеть, что выражение для $p(x, y)$ превращается в равенство

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad (2)$$

– совместная плотность величин X и Y равна произведению плотностей вероятности

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-M(X))^2}{2\sigma_X^2}}, \quad p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-M(Y))^2}{2\sigma_Y^2}}$$

для этих величин. Этот же вывод можно сделать, если предположить только выполнение равенства

$$\rho = \text{cor}(X, Y) = 0$$

– т.е. некоррелированность случайных величин X и Y . В этом случае (2) также выполняется: некоррелированность нормально распределенных случайных величин X и Y влечет их независимость. Напомним, что для случая общих случайных величин это утверждение вообще говоря неверно.

Пусть теперь случайные величины X и Y коррелированы. Поскольку плотность вероятности их совместного распределения задается формулой (1), то плотность для случайной величины X находится интегрированием:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy.$$

Прежде чем записывать это равенство более подробно сделаем замену переменной

$$v = \frac{y - M(Y)}{\sigma(Y)} \Rightarrow dy = \sigma(Y) dv$$

и обозначим

$$u = \frac{x - M(X)}{\sigma(X)}.$$

Тогда

$$\begin{aligned} p_X(x) &= \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho\cdot uv)} dv = \\ &= \frac{e^{-\frac{u^2}{2}}}{2\pi\sigma_X\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}(v-\rho\cdot u)^2} dv = \frac{e^{-\frac{u^2}{2}}}{2\pi\sigma_X} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-M(X))^2}{2\sigma_X^2}}. \end{aligned}$$

Из симметрии следует, что распределение Y также нормально с параметрами $(M(Y), \sigma^2(Y))$.

Рассмотрим далее условные плотности случайных величин X и Y . Согласно определениям

$$p(y|x) = \frac{p(x, y)}{p_X(x)}.$$

Подставляя выражения для $p(x, y)$ и $p_X(x)$ в эту дробь и выполняя несложные арифметические преобразования, можно проверить, что условное распределение случайной величины Y (при фиксированном значении случайной величины $X = x$) является нормальным с параметрами

$$M(Y|x) = M(Y) + \rho \frac{\sigma(Y)}{\sigma(X)}(x - M(X)), \quad D(Y|x) = D(Y) \cdot (1 - \rho^2).$$

Таким образом линия нормальной регрессии Y по X является прямой линией, условная дисперсия и вовсе постоянна, т.е. не зависит от значения x , принятого величиной X .

Лекция 10.

Закон больших чисел. Теоремы Бернулли и Чебышева. Центральная предельная теорема Ляпунова.

Как уже говорилось, характеристика $\sigma_X = \sqrt{D(X)}$ представляет из себя некоторую среднюю меру отклонений $(X - M(X))$ случайной величины X от математического ожидания $M(X)$. Поэтому естественно ожидать, что отклонения, существенно превышающие по абсолютной величине σ_X , должны быть маловероятны. В случае нормального распределения с параметрами (ν, σ) вероятность

$$Q(t) = P(|X - \nu| \geq t \cdot \sigma), \quad t > 0, \quad (1)$$

изображается площадью под нормальной кривой вне интервала $(-t, t)$. Для $t = 3$ эта вероятность составляет всего 0.0027, при $t = 4$ она уменьшается до 0.000063, а при $t = 6$ приблизительно равна $2 \cdot 10^{-9}$.

Оказывается, что свойство стремления к нулю величины $Q(t)$ из (1) свойственно не только случайным величинам X , распределенным по нормальному закону, но и общим случайным величинам X , обладающим моментами двух первых порядков. Именно, пусть X — произвольная случайная величина, $\nu = M(X)$, $\sigma = \sqrt{D(X)}$. Тогда при любом $t > 0$ выполняется неравенство (называемое неравенством Чебышева):

$$Q(t) = P(|X - \nu| \geq t \cdot \sigma) \leq \frac{1}{t^2}. \quad (2)$$

Для доказательства рассмотрим сначала случайную величину Z , принимающую лишь неотрицательные значения. Тогда имеет место неравенство

$$P(Z \geq \tau) \leq \frac{M(Z)}{\tau}. \quad (3)$$

Для определенности предположим, что величина Z непрерывно распределена с плотностью вероятности $p(z)$. По условию $p(z) = 0$ при $z < 0$ и для $\tau > 0$

$$P(Z \geq \tau) = \int_{\tau}^{\infty} p(z) dz.$$

Кроме того,

$$M(Z) = \int_0^{\infty} zp(z) dz = \int_0^{\tau} zp(z) dz + \int_{\tau}^{\infty} zp(z) dz \geq \int_{\tau}^{\infty} zp(z) dz \geq \tau \int_{\tau}^{\infty} p(z) dz = \tau P(Z \geq \tau),$$

что и доказывает (3). Для дискретных случайных величин схема доказательства (3) остается такой же, только интегралы заменяются суммами. Покажем теперь как из (3) следует (2). Напомним, что нам задана произвольная случайная величина X , для которой ν, σ – ее математическое ожидание и среднее квадратичное отклонение, $t > 0$ – произвольно. Пусть $Z = (X - \nu)^2$, $\tau = \epsilon^2$ ($\epsilon > 0$). Воспользуемся тем, что неравенство $(X - \nu)^2 \geq \epsilon^2$ эквивалентно неравенству $|X - \nu| \geq \epsilon$, и выпишем (3) для так определенных Z и τ :

$$P(|X - \nu| \geq \epsilon) = P((X - \nu)^2 \geq \epsilon^2) = P(Z \geq \tau) \leq \frac{M(Z)}{\tau} = \frac{M((X - \nu)^2)}{\epsilon^2}$$

или

$$P(|X - \nu| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}. \quad (4)$$

Положим $\epsilon = t \cdot \sigma$. Тогда

$$P(|X - \nu| \geq t \cdot \sigma) \leq \frac{D(X)}{t^2 \cdot \sigma^2} = \frac{1}{t^2}.$$

Теперь рассмотрим две фундаментальные теоремы теории вероятностей, имеющие обширный круг приложений. Обе они относятся к закону распределения числа появлений случайного события в данной серии независимых испытаний, т.е. последовательности наблюдений независимых испытаний одной и той же случайной величины X . Заметим прежде всего, что суммарное число появлений события в n независимых испытаниях можно рассматривать как сумму n независимых величин X_s , $s = 1, \dots, n$, каждая из которых принимает значение 1 или 0 в зависимости от того произошло или нет событие, связанное с X , в s -ом испытании: сумма

$$S_n = X_1 + X_2 + \dots + X_n$$

и есть число m появлений события в серии испытаний. Пусть p – вероятность появления события, связанного с исходной случайной величиной X , $q = 1 - p$ – вероятность противоположного события. Тогда

$$M(X_s) = q \cdot 0 + p \cdot 1 = p, \quad D(X_s) = M(X_s^2) - M(X_s)^2 = p - p^2 = pq, \quad s = 1, \dots, n.$$

Поэтому, в силу независимости испытаний,

$$M\left(\frac{S_n}{n}\right) = p, \quad D\left(\frac{S_n}{n}\right) = \frac{npq}{n^2} = \frac{pq}{n}.$$

Таким образом, с ростом числа испытаний n наблюдается убывание к нулю среднее квадратичное отклонение среднего арифметического независимых величин $\frac{S_n}{n}$ от среднего значения этой суммы. По неравенству Чебышева при любом фиксированном $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) < \frac{pq}{n\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty.$$

Полученное утверждение носит название теоремы Бернулли и гласит, что вероятность того, что среднее арифметическое большого числа независимых одинаково распределенных двузначных случайных величин будет отличаться от своего математического ожидания как минимум на некоторую фиксированную положительную величину, стремится к нулю с ростом числа испытаний.

Факт устойчивости средних арифметических большого числа одинаково распределенных независимых величин к их общему (в силу одинаковой распределенности) математическому ожиданию, имеет место и в более общей ситуации: пусть X_s – последовательность независимых одинаково распределенных случайных величин (т.е., в частности, $M(X_s) = \nu$, $D(X_s) = \sigma^2$). Тогда, так же как и выше, для

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

выполняются

$$M(\bar{X}_n) = \nu, \quad D(\bar{X}_n) = \frac{\sigma^2}{n}$$

и по неравенству Чебышева (4)

$$P(|\bar{X}_n - \nu| > \varepsilon) < \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0,$$

когда n стремится к бесконечности, а $\varepsilon > 0$ фиксировано. Таким образом, осредняя достаточно большое число независимых одинаково распределенных случайных величин, мы получим с вероятностью, как угодно близкой к единице, значение, сколь угодно мало отличающееся от общего математического ожидания величин. Это предложение, составляющее важный частный случай ”закона больших чисел”, было установлено П.Л. Чебышевым. Более общий случай этой теоремы выглядит так

Теорема (”закон больших чисел”). Пусть $X_1, X_2, \dots$ – последовательность попарно некоррелированных случайных величин, $M(X_n) = \nu$, $D(X_n) = \sigma^2 < \infty$, $n = 1, 2, \dots$. Тогда для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + \dots + X_n}{n} - \nu\right| > \varepsilon\right) = 0.$$

Многие задачи теории вероятностей и математической статистики связаны с изучением суммы независимых величин. Так теоремы, относящиеся к закону больших чисел, имеют дело с такими суммами. Основной задачей при этом, однако, являются нахождение и изучение поведения закона распределения суммы большого числа независимых слагаемых. Пусть сначала X и Y – две дискретные независимые величины одного испытания и $Z = X + Y$. Возможное значение z величины Z всегда представляет собой сумму $z = x + y$ двух возможных значений слагаемых x и y случайных величин X и Y . По правилу сложения имеем

$$P(Z = z) = \sum_{x+y=z} P(X = x, Y = y),$$

где суммирование распространяется на те пары возможных значений x и y , которые в сумме дают z . Но в силу независимости X и Y

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y).$$

Поэтому

$$P(Z = z) = \sum_{x+y=z} P(X = x) \cdot P(Y = y) = \sum_x' P(X = x) \cdot P(Y = z - x),$$

причем последняя сумма $\sum_x'$ распространена не на все значения x , а только на такие, для которых $z - x$ равно одному из возможных значений Y . Условимся полагать $P(Y = z - x)$ равной нулю всякий раз, когда $z - x$ не принадлежит к числу возможных значений Y . Тогда

$$P(Z = z) = \sum_x P(X = x) \cdot P(Y = z - x)$$

по всем значениям x . Точно также получается равенство

$$P(Z = z) = \sum_y P(Y = y) \cdot P(X = z - y).$$

Эти два равенства для функций распределения в дискретном случае, используя чуть иные обозначения, можно записать так

$$P_Z(z) = \sum_x P_X(x) \cdot P_Y(z - x), \quad P_Z(z) = \sum_y P_Y(y) \cdot P_X(z - y). \quad (5)$$

Аналогично разбирается случай суммы двух независимых непрерывных случайных величин с плотностями вероятностей $p_X(x)$ и $p_Y(y)$ соответственно. Можно проверить, что для этого случая плотность вероятности суммы $Z = X + Y$ случайных величин, находится по формулам, похожим на (5):

$$p_Z(z) = \int_{-\infty}^{+\infty} p_X(x) \cdot p_Y(z - x) dx, \quad p_Z(z) = \int_{-\infty}^{+\infty} p_Y(y) \cdot p_X(z - y) dy.$$

Отыскание и исследование распределения, представляющего сумму многих случайных величин, облегчается применением так называемых производящих функций этих случайных величин. Производящей функцией случайной величины X называют

$$m_X(t) = M(e^{X \cdot t}),$$

где $t \in \mathbb{R}$ – вспомогательный параметр. В частности для дискретной случайной величины X с функцией вероятности $p(x)$

$$m_X(t) = \sum_x e^{xt} p(x).$$

При $t = 0$ получим

$$m_X(0) = \sum_x e^{x \cdot 0} p(x) = 1.$$

Продифференцируем несколько раз производящую функцию (в предположении абсолютной сходимости получающихся рядов) и подставим $t = 0$:

$$m_X^{(k)}(0) = \sum_x x^k e^{x \cdot 0} p(x) = \sum_x x^k p(x) = M(X^k).$$

При определенных ограничениях эти формулы справедливы также и для непрерывных случайных величин. Таким образом, если известна производящая функция некоторой случайной величины X , то ее первые моменты могут быть найдены по формулам $M(X^k) = m_X^{(k)}(0)$. Полученные равенства означают еще, что разложение в степенной ряд производящей функции случайной величины X имеет вид

$$m_X(t) = 1 + \frac{\nu_1 t}{1!} + \frac{\nu_2 t^2}{2!} + \dots + \frac{\nu_k t^k}{k!} + \dots, \quad \nu_k = M(X^k), \quad k \in \mathbb{N}.$$

Из определения производящей функции следует, что если $m_X(t)$ – производящая функция случайной величины X , то для случайной величины $Y = aX + b$, $a, b \in \mathbb{R}$,

$$m_Y(t) = M(e^{(aX+b)t}) = M(e^{X \cdot at} \cdot e^{bt}) = e^{bt} m_X(at).$$

Отметим без доказательства два основных положения общей теории производящих функций. Первое из них называется теоремой единственности и утверждает, что производящая функция однозначно определяет распределение вероятностей, так что не только каждому закону отвечает определенная производящая функция, но и наоборот, каждой производящей функции соответствует единственное распределение вероятностей. Второе положение – теорема ”непрерывности”, утверждающая, что если последовательность производящих функций $m_{X_i}(t)$, построенных для случайных величин X_i с функциями распределения вероятности $p_{X_i}(x)$, в каждой точке $t \in \mathbb{R}$ сходится к производящей функции $m_X(t)$ случайной величины X с функцией распределения вероятности $p_X(x)$, то и сами законы $p_{X_i}(x)$ сходятся к закону распределения вероятности $p_X(x)$.

Вычислим для примера производящую функцию нормальной с параметрами (a, σ^2) случайной величины:

$$m(t) = M(e^{Xt}) = \frac{1}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{xt - \frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(z\sigma+a)t - \frac{z^2}{2}} dz = e^{at + \frac{t^2\sigma^2}{2}}.$$

Теорема. Производящая функция для суммы S независимых случайных величин $X_1, X_2, \dots, X_n$ равна произведению производящих функций компонент.

Действительно, если $S = X_1 + \dots + X_n$, то, из-за независимости величин $X_1, X_2, \dots, X_n$,

$$m_S(t) = M(e^{(X_1 + \dots + X_n) \cdot t}) = M(e^{X_1 \cdot t} \cdot \dots \cdot e^{X_n \cdot t}) = M(e^{X_1 \cdot t}) \cdot \dots \cdot M(e^{X_n \cdot t}) = m_{X_1}(t) \cdot \dots \cdot m_{X_n}(t).$$

В частности, если все величины X_i распределены одинаково ($m_{X_i}(t) = m(t)$), то полученное равенство превращается в

$$m_S(t) = (m(t))^n.$$

Предположим теперь, что для независимых одинаково распределенных случайных величин X_i

$$M(X_i) = 0, \quad M(X_i^2) = \sigma^2, \quad i = 1, 2, \dots$$

определены их производящие функции. Тогда для суммы $S = \sum_{i=1}^n X_i$ этих случайных величин

$$M(S) = 0, \quad D(S) = n\sigma^2.$$

Нормируем случайную величину S :

$$\bar{S} = \frac{S}{\sigma\sqrt{n}}, \quad M(\bar{S}) = 0, \quad D(\bar{S}) = 1.$$

Тогда

$$m_{\bar{S}}(t) = \left(m\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n.$$

По формуле Тейлора

$$e^u = 1 + u + u^2/2 + u^3/6e^{\theta u}, \quad \theta \in (0, 1)$$

и, значит,

$$m\left(\frac{t}{\sigma\sqrt{n}}\right) = M\left(1 + \frac{X_i t}{\sigma\sqrt{n}} + \frac{X_i^2 t^2}{2\sigma^2 n} + \frac{X_i^3 t^3}{6\sigma^3 \sqrt{n}^3} e^{\theta \frac{X_i t}{\sigma\sqrt{n}}}\right) = 1 + \frac{t^2}{2n} + O(n^{-3/2}).$$

Оказывается, что при определенных достаточно общих ограничениях при вычислении предела

$$\lim_{n \rightarrow \infty} \left(m\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n = \lim_{n \rightarrow \infty} \left(1 + \frac{t^2}{2n} + O(n^{-3/2}) \right)^n$$

можно пренебречь слагаемым $O(n^{-3/2})$ в выражении справа (оно стремится к нулю с ростом n значительно быстрее, чем $\frac{t^2}{2n}$). Поэтому

$$\lim_{n \rightarrow \infty} m_{\bar{S}}(t) = \lim_{n \rightarrow \infty} \left(1 + \frac{t^2}{2n} \right)^n = e^{t^2/2}.$$

На основании теоремы непрерывности отсюда следует, что закон распределения $P_{\bar{S}}(x)$ нормированной суммы $\bar{S}$ сходится к нормальному закону распределения с параметрами $(0, 1)$ или, как еще говорят, эта нормированная сумма асимптотически нормальна. Это означает, в частности, что каков бы ни был исходный закон распределения независимых случайных величин X_i , после усреднения при достаточно большом n закон распределения случайной величины $\bar{S}$ можно приблизительно считать равным стандартному нормальному закону. В этом и состоит смысл утверждений, называемых центральными предельными теоремами. В заключение приведем пример формулировки одной из таких теорем:

Теорема. Пусть $X_1, \dots, X_n, \dots$ – независимые одинаково распределенные случайные величины с конечными математическим ожиданием a и дисперсией σ^2 . Тогда для любого $x \in \mathbb{R}$

$$P\left(\frac{X_1 + X_2 + \dots + X_n - na}{\sigma\sqrt{n}} < x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{v^2}{2}} dv$$

при $n \rightarrow \infty$.

Лекция 11.

Основные понятия математической статистики. Генеральная совокупность и выборка. Вариационный ряд, статистический ряд. Группированная выборка. Группированный статистический ряд. Выборочная функция распределения и гистограмма. Числовые характеристики статистического распределения: выборочное среднее, оценки дисперсии, начальных и центральных моментов. Некоторые свойства статистических оценок параметров распределения: несмещенность, состоятельность.

Основной задачей математической статистики является разработка методов получения обоснованных выводов о массовых явлениях и процессах из данных наблюдений или экспериментов. Эти выводы и заключения относятся не к отдельным испытаниям, из повторения которых и складывается данное массовое явление, а представляют утверждения об общих вероятностных характеристиках данного процесса, т.е. о вероятностях, законах распределения, математических ожиданиях и т.д. Такое использование фактических данных как раз и является отличительной чертой статистического метода.

Таким образом, в математической статистике исследуются способы получения выводов на основе эмпирических данных, взятых из наблюдений. Случайной выборкой объема n (или просто выборкой) называется случайный вектор:

$$x_1, x_2, \dots, x_n, \quad (1)$$

где x_i обычно предполагаются независимыми и имеющими одну и ту же функцию распределения $F(x) = P(x_i < x)$ случайными величинами. Случайная выборка (1), в этом случае, является математической моделью последовательно производимых измерений. В то же время эта совокупность результатов измерений обычно мыслится как представитель целого статистического ансамбля, который можно было бы иметь при многократном повторении системы опытов, принесший один раз результаты (1). При этом полный перечень результатов опытов (т.е. продолжение (1) в полном объеме), который на данный момент нам не доступен, обычно называют генеральной совокупностью. Итак, $x_1, x_2, \dots, x_n$ — случайные величины с одинаковым и нам заранее неизвестным законом распределения $F(x) = P(x_i < x)$, $x \in \mathbb{R}$. Вычислим его приближенно.

Определение. Эмпирической (выборочной) функцией распределения называется функция $F_n(x)$, задаваемая соотношением

$$F_n(x) = \frac{\text{число } x_i < x}{n}.$$

Другим (эквивалентным приведенному) подходом определить функцию $F_n(x)$ является такой: выстроим значения (1) в порядке неубывания

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Получившийся ряд чисел $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ называют вариационным рядом (в частности,

$$x_{(1)} = \min\{x_1, x_2, \dots, x_n\}, \quad x_{(n)} = \max\{x_1, x_2, \dots, x_n\},$$

разность $x_{(n)} - x_{(1)} = w$ называется размахом выборки. Пусть по определению $x_{(0)} = -\infty$, $x_{(n+1)} = +\infty$. Эмпирической (выборочной) функцией распределения называется функция $F_n(x)$, равная $\frac{i}{n}$ при $x \in (x_{(i)}, x_{(i+1)}]$, $i = 0, 1, \dots, n$. Таким образом, $F_n(x)$ — неубывающая, кусочно-постоянная функция, равная нулю при $x \leq x_{(1)}$ и равная единице при $x > x_{(n)}$.

Продолжим разговор об эмпирической функции. Пусть выборка $x_1, x_2, \dots, x_n$ содержит ровно k различных чисел $z_1, z_2, \dots, z_k$, причем z_i встречается n_i раз ($i = 1, 2, \dots, k$). Число n_i называется частотой элемента выборки z_i . Очевидно, что $\sum n_i = n$. Статистическим рядом называется последовательность (z_i, n_i) . Обычно статистический ряд записывается в виде таблицы, первая строка которой содержит элементы z_i , а вторая — n_i частоты.

При большом объеме выборки ее элементы объединяют в группы, представляя результаты опытов в виде группированного статистического ряда. Для этого интервал (длины w), содержащий все элементы выборки, разбивается на некоторое количество l непересекающихся интервалов. Вычисления значительно упрощаются, если эти интервалы имеют одинаковую длину $b \approx w/l$. Во всем дальнейшем изложении рассматривается именно этот случай. После того как частичные интервалы выбраны, определяют частоты

– количество n_i элементов выборки, попавших в i -й интервал (элемент, совпадающий с верхней границей интервала, относится к последующему интервалу). Получающийся статистический ряд в верхней строке содержит середины z_i интервалов группировки, а в нижней – частоты n_i ($i = 1, 2, \dots, l$). Наряду с частотами одновременно подсчитываются также накопленные частоты $\sum_{i=1}^j n_i$, относительные частоты n_j/n и накопленные относительные частоты $\sum_{i=1}^j n_i/n$, $j = 1, 2, \dots, l$. Полученные результаты сводятся в таблицу, называемую таблицей частот группированной выборки.

Итак, еще раз повторим, что эмпирической функцией $F_n(x)$ называют, по сути, вероятностное распределение дискретной случайной величины, принимающей значения $x_1, x_2, \dots, x_n$ с вероятностями $1/n$. Она может быть определена по значениям накопленных частот соотношением

$$F_n(x) = \frac{1}{n} \sum_{z_i < x} n_i.$$

Теорема (Гливенко). Пусть $F_n(x)$ – эмпирическая функция распределения, построенная по выборке объема n из генеральной совокупности с функцией распределения $F(x)$. Тогда для любого $x \in (-\infty, +\infty)$ и любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| \leq \varepsilon) = 1.$$

Заметим, что $F_n(x)$ представимо в виде

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_x(x_i),$$

где функция $I_x(y)$ равна 0, если $x \leq y$, и равна 1, если $x > y$. Так как x_i – независимые одинаково распределенные случайные величины, то $I_x(x_i)$ – также независимы. При этом случайные величины $I_x(x_i)$ принимают только два значения 0 и 1, причем выполняется равенство

$$P(I_x(x_i) = 1) = P(x_i < x) = F(x),$$

означающее, что

$$M(I_x(x_i)) = F(x), \quad D(I_x(x_i)) = M(I_x^2(x_i)) - (M(I_x(x_i)))^2 = F(x) - F^2(x) = F(x)(1 - F(x)).$$

Воспользуемся "законом больших чисел":

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{I_x(x_1) + \dots + I_x(x_n)}{n} - F(x)\right| > \varepsilon\right) = \lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| > \varepsilon) = 0,$$

что и завершает доказательство теоремы.

Для наглядного представления выборки используют гистограмму и полигон частот. Гистограммой частот группированной выборки называется кусочно-постоянная функция, постоянная на интервалах группировки и принимающая на каждом из них значения n_i/b , $i = 1, 2, \dots, l$, соответственно. Площадь ступенчатой фигуры под графиком гистограммы равна объему выборки n . Аналогично определяется гистограмма относительных частот. Площадь соответствующей ступенчатой фигуры для нее равна единице. При увеличении объема выборки и уменьшении интервала группировки гистограмма относительных частот является статистическим аналогом плотности распределения $p_X(x)$ генеральной совокупности.

Пусть $x_1, x_2, \dots, x_n$ – выборка объема n из генеральной совокупности с функцией распределения $F(x)$. Числовые характеристики этого выборочного распределения называются выборочными (эмпирическими) числовыми характеристиками. Так, например, выборочным средним и выборочной дисперсией будут величины

$$M_X^* = \frac{1}{n} \sum x_i, \quad D_X^* = \frac{1}{n} \sum x_i^2 - (M_X^*)^2. \quad (2)$$

Моменты старших порядков, построенные для распределения $F_n(x)$, называют соответственно выборочными моментами, построенными по результатам наблюдений (1). Формулы (2) и их аналоги принято

называть оценками (т.е. правилами восстановления неизвестных параметров – в данном случае математического ожидания и старших моментов), построенными по результатам наблюдений (1) случайной величины X с функцией распределения $F(x)$.

Значения, полученные по формулам (2), и вообще любую характеристику, полученную на основании данных выборки, следует рассматривать, как значение некоторой случайной величины, варьирующей от выборки к выборке. Рассмотрим, например, величину $\bar{X}_n = M_X^*$ как среднее арифметическое выборки (1), состоящей из независимых величин с одним и тем же законом распределения вероятности $F(x)$ (имеющих, в частности, одинаковые математические ожидания и дисперсии, совпадающие с $M(X)$ и $D(X)$). Прежде мы уже неоднократно подсчитывали, что в этом случае

$$M(\bar{X}_n) = M(X), \quad D(\bar{X}_n) = \frac{D(X)}{n}. \quad (3)$$

Согласно закону больших чисел для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - M(X)| > \varepsilon) = \lim_{n \rightarrow \infty} P\left(\left|\frac{x_1 + \dots + x_n}{n} - M(X)\right| > \varepsilon\right) = 0. \quad (4)$$

Выборочные оценки того или иного параметра (в данном случае $\bar{X}_n$ для $M(X)$) называются состоятельными оценками этого параметра, если с ростом n размера выборки наблюдается, как говорят, сходимость по вероятности выборочной оценки к этому параметру (в данном случае это означает выполнение (4)).

С другой стороны выражение D_X^* (еще раз обратим внимание, что с одной стороны это есть случайная величина, зависящая от результатов выборки (1), а с другой – оценка для еще одного параметра наблюдаемой величины, – для ее дисперсии $D(X)$) может быть записано в следующем виде

$$\begin{aligned} D_X^* &= \frac{1}{n} \sum (x_i - M_X^*)^2 = \frac{1}{n} \sum (x_i - M(X) + M(X) - M_X^*)^2 = \frac{1}{n} \sum (x_i - M(X))^2 + \\ &+ \frac{2(M(X) - M_X^*)}{n} \sum (x_i - M(X)) + (M(X) - M_X^*)^2 = \frac{1}{n} \sum (x_i - M(X))^2 - (M_X^* - M(X))^2. \end{aligned}$$

Таким образом,

$$M(D_X^*) = M\left(\frac{1}{n} \sum (x_i - M(X))^2 - (M_X^* - M(X))^2\right) = \frac{n-1}{n} D(X).$$

Таким образом, в отличие от M_X^* , выборочная характеристика D_X^* имеет математическое ожидание, не совпадающее с оцениваемым параметром (т.е. с дисперсией $D(X)$) ни при каком значении объема выборки n . Чтобы исправить это несоответствие вместо выборочной дисперсии из (2) рассматривают уточненную выборочную дисперсию

$$\frac{n}{n-1} D_X^*. \quad (5)$$

Выборочные характеристики, обладающие тем свойством, что их математические ожидания при любом объеме выборки равны оцениваемому параметру, называются несмещенными оценками. Примерами несмещенных оценок служат величины M_X^* и $\frac{n}{n-1} D_X^*$.