МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное бюджетное государственное образовательное учреждение высшего образования

Донской государственный технический университет

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Курс лекций

Составитель: зав. кафедрой высшей математики, д.ф.-м.н., профессор Павлов И.В.

Рецензенты: к.ф.-м.н., доцент А.М. Можаев,
к.ф.-м.н., доцент Г.А. Власков

Ростов-на-Дону

2022

ЛЕКЦИИ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

ЧАСТЬ 1: ПЕРВОНАЧАЛЬНЫЕ ОПРЕДЕЛЕНИЯ И ФАКТЫ

Лекция 1

Относительно того, что твердо известно и не подлежит сомнению, мы говорим, что ***знаем*** *и* ***понимаем****; относительно всего прочего – что только догадываемся или* ***предполагаем****. Делать о какой-либо вещи предположения – все равно, что измерять ее вероятность. Поэтому искусство предположений мы определяем как искусство возможно точнее измерять вероятности вещей затем, чтобы в наших суждениях или действиях мы могли выбирать или следовать тому, что будет найдено лучшим, более удовлетворительным, спокойным и разумным.*

Я. Бернулли*,*
Базель, 1713 г.

Относительная частота и классическое определение вероятности

Основным понятием теории вероятностей является понятие случайного события, то есть события, которое (при некоторых обстоятельствах) может совершиться, а может и не совершиться. Например, выпадение орла при однократном бросании монеты есть случайное событие; попадание в цель при стрельбе есть также случайное событие и т.д.

Предположим, что мы совершаем серию одинаковых испытаний (например, многократно бросаем монету). Зафиксируем некоторое событие A, которое может реализовываться в результате каждого из этих испытаний (например, A – событие, состоящее в выпадении орла). Пусть n^* – количество произведенных испытаний, а m^* – число реализаций события A. Относительной частотой появления события A будем называть отношение $p^* = \frac{m^*}{n^*}$. Если мы совершим еще одну серию таких же испытаний, причем новое n^* будет приблизительно таким же, как и прежнее, то новая относительная частота p^* может существенно отличаться от старой. Однако в результате многочисленных экспериментов было сделано заключение, что существует число p, такое, что относительные частоты появления события A при большом числе повторяющихся испытаний лишь незначительно отличаются от этого числа p. Данный эмпирический факт символически записывается следующим образом:

$$\frac{m^*}{n^*} \to p \text{ при } n^* \to \infty. \qquad (1)$$

Число p называется вероятностью реализации события A, что символически записывается в виде: $P(A) = p$.

Вероятность p есть объективная характеристика возможности реализации события A в данном испытании. При большом числе испытаний относительная частота мало отличается от вероятности p (исключения могут составлять весьма редкие серии испытаний, которыми можно пренебречь). Например, при многократном бросании монеты мы в большинстве случаев получим относительную частоту выпадения орла, мало отличающуюся от $p = \frac{1}{2}$.

Соотношение (1) обычно формулируют следующим образом: когда число n^* повторяющихся испытаний бесконечно растет, относительная частота события A стремится к вероятности p реализации события A. Это нестрогая формулировка хорошо известной теоремы Я. Бернулли, являющейся краеугольным камнем современной теории вероятностей.

Пусть теперь мы находимся в рамках некоторого испытания. Придавая каждому событию A, связанному с данным испытанием, некоторую вероятность $P(A)$, мы тем самым создаем математическую модель, описывающую (по нашим понятиям, адекватным образом) данное испытание. Построив математическую модель, мы имеем возможность для решения поставленных задач пустить в ход весьма развитый математический аппарат теории вероятностей, в основе которого лежит аксиоматика, впервые предложенная академиком А.Н.Колмогоровым в 1929 г. Проиллюстрируем процесс построения математической модели на следующем примере.

Пример 1. Рассмотрим эксперимент, состоящий в однократном бросании на идеальную горизонтальную поверхность игральной кости (шестигранного кубика), на гранях которой нанесены цифры от 1 до 6. Мы считаем, что игральная кость также имеет идеальную конфигурацию. Ясно, что в результате бросания мы можем получить шесть исходов (или, говоря математическим языком, шесть ***элементарных событий***), которые мы обозначим через $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$, где ω_k – событие, заключающееся в выпадении числа k (k=1,2,...,6). Ввиду симметрии игральной кости, введенные элементарные события можно считать равновероятными. Действительно, если мы совершим большую серию из n^* бросаний игральной кости, то каждое ω_k совершится приблизительно $\frac{n^*}{6}$ раз, то есть относительная частота реализации этого события будет приблизительно равна $\frac{1}{6}$. Это говорит о том, что каждому событию ω_k нужно придать вероятность $\frac{1}{6}$, то есть положить $P(\omega_k) = \frac{1}{6}$. На этом фактически завершается процесс построения математической модели данного эксперимента. Эта математическая модель называется ***вероятностным пространством***.

□

Определение 1. ***Классическим вероятностным пространством*** называется тройка $(\Omega, \mathcal{F}, P)$, где $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ – некоторое множество, состоящее из конечного числа элементов; $\mathcal{F}$ – совокупность всех подмножеств множества Ω (включая само Ω и пустое множество $\varnothing$); P – функция, определенная на $\mathcal{F}$ следующим образом: если подмножество $A \in \mathcal{F}$ состоит из m элементов, то $P(A) := \frac{m}{n}$. При этом элементы ω_k называются ***элементарными событиями***, множество Ω – ***пространством элементарных событий***, множество A – ***событием***, событие Ω – ***достоверным событием***, событие $\varnothing$ – ***невозможным событием***, совокупность событий $\mathcal{F}$ – ***алгеброй событий***, функция P – ***вероятностью***.

□

Условимся число элементарных событий, из которых состоит событие *A* (и которые по этой причине называются ***благоприятствующими*** событию *A*), обозначать через |*A*|. Из определения 1 получаем: $P(A)=\frac{|A|}{n}$, $P(\Omega)=\frac{|\Omega|}{n}=\frac{n}{n}=1$, $P(\varnothing)=\frac{|\varnothing|}{n}=\frac{0}{n}=0$, $P(\omega_k)=\frac{|\omega_k|}{n}=\frac{1}{n}$, где $k=1,2,...,n$. Из последнего равенства, в частности, следует, что математическую модель, построенную в примере 1, можно считать классическим вероятностным пространством.

Пример 1 (продолжение). Пусть A_2 (соотв. A_3) – событие, состоящее в том, что выпавшее в результате бросания кости число делится на 2 (соотв., на 3). Вычислим вероятности событий A_2 и A_3. Так как $A_2=\{\omega_2,\omega_4,\omega_6\}$, а $A_3=\{\omega_3,\omega_6\}$, то применяя классическое определение вероятности, имеем:

$$P(A_2)=\frac{|A_2|}{n}=\frac{3}{6}=\frac{1}{2}, \qquad P(A_3)=\frac{|A_3|}{n}=\frac{2}{6}=\frac{1}{3}.$$

□

Пример 2. Из колоды, содержащей 36 карт, наудачу выбирается одна карта. Найти вероятность того, что эта карта бубновой масти.

Считая, что вероятность изъятия из колоды любой карты равна $\frac{1}{36}$, можно применить классическое определение вероятности с $n=36$. Пусть *A* – событие, состоящее в том, что выбрана карта бубновой масти. Так как всего в колоде 9 бубновых карт, то $|A|=9$. Поэтому $P(A)=\frac{9}{36}=\frac{1}{4}$.

□

Пример 3. Одновременно бросаются две монеты. Какова вероятность того, что на обеих монетах выпадет решка?

Предположив, что результаты бросаний равновероятны, применим классическую схему. Тогда пространство элементарных событий Ω можно рассматривать как совокупность пар (орел, решка), (орел, орел), (решка, решка), (решка, орел), то есть $n=|\Omega|=4$. Ясно, что вероятность элементарного события (решка, решка) равна $\frac{1}{4}$.

□

Пример 4. Партия из 100 изделий содержит 10 бракованных изделий. Какова вероятность того, что среди выбранных наудачу пяти изделий ровно два изделия окажутся бракованными?

Предполагая, что каждая выборка пяти изделий из 100 равновероятна любой другой такой же выборке, применим классическую схему. Как известно, число качественно различных выборок (или, что то же самое, число сочетаний) из *r* элементов по *k* равно:

$$C_r^k=\frac{r!}{k!\cdot(r-k)!}. \tag{2}$$

Поэтому в нашем случае $n=|\Omega|=C_{100}^5$. Пусть *A* – событие, состоящее из выборок, содержащих два бракованных изделия и три качественных. Так как два бракованных изделия можно изъять только из 10-ти бракованных, три качественных изделия – из 90 качественных и каждая бракованная пара может при выборе состыковаться с каждой качественной тройкой, то число элементарных событий (пятерок изделий), благоприятствующих событию *A* равно $|A|=C_{10}^2\cdot C_{90}^3$. Таким образом,

$$P(A)=\frac{|A|}{n}=\frac{C_{10}^{2}\cdot C_{90}^{3}}{C_{100}^{5}}=\frac{5!\cdot 95!\cdot 10!\cdot 90!}{100!\cdot 2!\cdot 8!\cdot 3!\cdot 87!}=\frac{3!\cdot 4\cdot 5\cdot 95!\cdot 8!\cdot 9\cdot 10\cdot 87!\cdot 88\cdot 89\cdot 90}{95!\cdot 96\cdot 97\cdot 98\cdot 99\cdot 100\cdot 2!\cdot 8!\cdot 3!\cdot 87!}=$$

$$=\frac{4\cdot 5\cdot 9\cdot 10\cdot 88\cdot 89\cdot 90}{96\cdot 97\cdot 98\cdot 99\cdot 100\cdot 2}=\frac{1335}{19012}\approx 0{,}07\,.$$

□

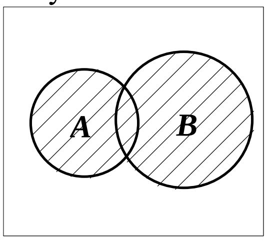
Действия над событиями. Алгебра событий

В дальнейшем мы будем рассматривать не только классические вероятностные пространства, где, в частности, Ω конечно, но и такие, у которых Ω может быть произвольным множеством. В этом последнем случае на совокупность событий F (см. определение 1) нужно накладывать дополнительные требования, так как схема

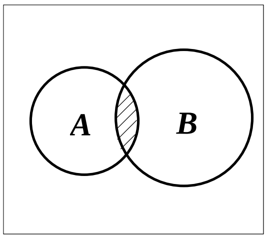
«F = совокупность ***всех*** подмножеств Ω»

не дает возможностей для продуктивного развития теории.

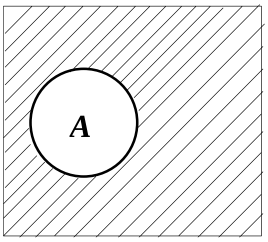
Итак, пусть Ω – произвольное множество (произвольное пространство элементарных событий), а F – какая-то совокупность подмножеств (событий) на Ω. Для определенности Ω мы будем изображать квадратом, а события из F – кругами в этом квадрате. Пусть A и B – события из F.



Суммой событий A и B называется событие $A+B$, состоящее из элементарных событий, благоприятствующих по крайней мере одному из событий A или B. Это означает, что событие $A+B$ совершается тогда и только тогда, когда совершается хотя бы одно из событий A или B. Геометрически сумме событий соответствует операция объединения множеств.



Произведением событий A и B называется событие $A\cdot B$, состоящее из элементарных событий, благоприятствующих одновременно и событию A, и событию B. Это означает, что событие $A\cdot B$ совершается тогда и только тогда, когда совершаются оба события A и B. Геометрически произведению событий соответствует операция пересечения множеств.



Противоположным событию A называется событие $\overline{A}$, состоящее из элементарных событий, не благоприятствующих событию A. Это означает, что событие $\overline{A}$ совершается тогда и только тогда, когда не совершается событие A. Геометрически противоположному событию соответствует операция дополнения множества.

Операции над событиями подчиняются точно тем же законам, что и операции над множествами. Например, справедливы соотношения (законы Де Моргана):

$$\overline{A+B}=\overline{A}\cdot\overline{B},\quad \overline{A\cdot B}=\overline{A}+\overline{B}\,.$$

Читателю рекомендуется доказать эти соотношения.

<u>*Определение 2*</u>. Совокупность F подмножеств множества Ω называется ***алгеброй событий***, если $\Omega\in F$ и если $A+B\in F$, $A\cdot B\in F$, $\overline{A}\in F$ для любых $A\in F$ и $B\in F$.

□

Так как $\overline{\Omega}=\varnothing$, то из определения 2 немедленно следует, что $\varnothing\in F$ для любой алгебры событий F.

Пример 5. Пусть $\Omega = R$ – множество всех действительных чисел. Обозначим через F совокупность конечных объединений интервалов вида $[a,b)$, которые мы будем интерпретировать как события, состоящие в попадании точки в этот интервал (если a или b бесконечны, то этот интервал также считается входящим в F). Читателю предоставляется возможность доказать, что F – алгебра событий.

□

ЧАСТЬ 1: ПЕРВОНАЧАЛЬНЫЕ ОПРЕДЕЛЕНИЯ И ФАКТЫ

Лекция 2

Общее определение вероятностного пространства

Из определения 2 вытекает, что если алгебра событий F содержит конечную совокупность событий $A_1, A_2, \ldots, A_n$, то она содержит и события $\bigcup_{k=1}^{n} A_k := A_1 + A_2 + \ldots + A_n$ и $\bigcap_{k=1}^{n} A_k := A_1 \cdot A_2 \cdot \ldots \cdot A_n$. Однако, вообще говоря, неверно следующее утверждение: если алгебра событий F содержит бесконечную последовательность событий $A_1, A_2, \ldots, A_n, A_{n+1}, \ldots$, то она содержит события $\bigcup_{k=1}^{\infty} A_k := A_1 + A_2 + \ldots + A_n + A_{n+1} + \ldots$(которое выполняется, если выполняется хотя бы одно из событий $A_1, A_2, \ldots, A_n, A_{n+1}, \ldots$) и $\bigcap_{k=1}^{\infty} A_k := A_1 \cdot A_2 \cdot \ldots \cdot A_n \cdot A_{n+1} \cdot \ldots$(которое выполняется, если выполняется каждое из событий $A_1, A_2, \ldots, A_n, A_{n+1}, \ldots$).

Пример 5 (продолжение). Пусть $A_k = \left[\frac{1}{k}, 1\right)$. Тогда $\bigcup_{k=1}^{\infty} A_k = (0,1)$, однако ясно, что открытый интервал $(0,1)$ не входит в алгебру F. Аналогично, если $A_k = \left[0, \frac{1}{k}\right)$, то $\bigcap_{k=1}^{\infty} A_k = \{0\}$, однако одноточечное множество $\{0\}$ не входит в F.

□

Определение 3. Алгебра F называется σ-алгеброй, если из того, что $A_k \in \mathrm{F}$ при всех k=1,2,3,..., следует, что $\bigcup_{k=1}^{\infty} A_k \in \mathrm{F}$.

□

Из легко доказываемого соотношения $\overline{\bigcap_{k=1}^{\infty} A_k} = \bigcup_{k=1}^{\infty} \overline{A_k}$ следует, что если F – σ-алгебра и $A_k \in \mathrm{F}$ при всех k=1,2,3,..., то $\bigcap_{k=1}^{\infty} A_k \in \mathrm{F}$.

Можно доказать (доказательство не входит в нашу программу), что для любой алгебры событий F существует наименьшая σ-алгебра $\hat{\mathrm{F}}$, содержащая F.

Пример 5 (продолжение). В условиях данного примера σ-алгебра $\hat{\mathrm{F}}$ называется борелевской σ-алгеброй на $\Omega = R$. Структура $\hat{\mathrm{F}}$ достаточно сложна. В частности, $\hat{\mathrm{F}}$ содержит все открытые интервалы, одноточечные множества и т.п. Однако можно доказать, что существуют подмножества R, не принадлежащие $\hat{\mathrm{F}}$ (эти подмножества достаточно экзотичны и мы никогда не будем иметь с ними дела).

□

Определение 4. ***Вероятностным пространством*** называется тройка (Ω, F, P), где Ω – произвольное множество, F – σ-алгебра подмножеств множества Ω, P – функция, определенная на F и удовлетворяющая следующим условиям:

1. (Аксиома положительности). Для любого $A \in \mathcal{F}$ $\boxed{P(A) \ge 0}$, то есть вероятность любого события неотрицательна.

2. (Аксиома нормировки). $\boxed{P(\Omega)=1}$, то есть вероятность достоверного события равна единице.

3. (Аксиома σ-аддитивности). Если содержащаяся в $\mathcal{F}$ последовательность событий $A_1, A_2, ..., A_n, A_{n+1}, ...$ такова, что $A_i \cdot A_j = \varnothing$ при всех $i \neq j$, то $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) := \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k)$.

□

Заметим, что термины, связанные с общим определением вероятностного пространства, совпадают с терминами, данными в определении 1. Далее в этом пункте мы будем работать в некотором фиксированном вероятностном пространстве $(\Omega, \mathcal{F}, P)$.

<u>*Определение 5*</u>. События *A* и *B* называются ***несовместными***, если вероятность их произведения равна нулю, то есть если $P(A \cdot B)=0$.

□

Свойства вероятности

1. $\boxed{P(\overline{A})=1-P(A)}$.

<u>*Доказательство*</u>. Так как $A+\overline{A}=\Omega$ и $A \cdot \overline{A}=\varnothing$, то применяя аксиому 3 вероятности (см. определение 4), получаем $P(A)+P(\overline{A})=P(\Omega)$, откуда по аксиоме нормировки $P(A)+P(\overline{A})=1$, то есть $P(\overline{A})=1-P(A)$.

□

<u>*Следствие*</u>. Так как $\varnothing=\overline{\Omega}$, то $P(\varnothing)=P(\overline{\Omega})=`1-P(\Omega)=1-1=0$, то есть $\boxed{P(\varnothing)=0}$ (вероятность невозможного события равна нулю). Отсюда в свою очередь следует, что если $A \cdot B=\varnothing$, то события *A* и *B* несовместны (в литературе часто именно соотношение $A \cdot B=\varnothing$ принимается за определение несовместности событий).

2. $\boxed{P(A+B)=P(A)+P(B)-P(A \cdot B)}$.

<u>*Доказательство*</u>. Справедливо равенство $A+B=A+\overline{A} \cdot B$ (действительно, $\omega \in A+B$ равносильно тому, что либо $\omega \in A$, либо $\omega \in B$, но не входит в *A*). Так как $A \cdot (\overline{A} \cdot B) = (A \cdot \overline{A}) \cdot B = \varnothing \cdot B = \varnothing$, то по аксиоме 3 вероятности имеем: $P(A+B)=P(A+\overline{A} \cdot B)=P(A)+P(\overline{A} \cdot B)$. С другой стороны, так как $B=A \cdot B+\overline{A} \cdot B$ и $(A \cdot B) \cdot (\overline{A} \cdot B)=(B \cdot A) \cdot (\overline{A} \cdot B)=B \cdot (A \cdot \overline{A}) \cdot B=B \cdot \varnothing \cdot B=\varnothing$, то по той же аксиоме 3 $P(B)=P(A \cdot B)+P(\overline{A} \cdot B)$, откуда $P(\overline{A} \cdot B)=P(B)-P(A \cdot B)$. Подставляя это в ранее полученное выражение для $P(A+B)$, получаем требуемое.

□

<u>*Следствие*</u>. Для любых событий *A* и *B* $\boxed{P(A+B) \le P(A)+P(B)}$. Если события *A* и *B* несовместны, то справедлива формула: $P(A+B)=P(A)+P(B)$.

3. Если из события *A* следует событие *B* (то есть если $A \subset B$), то $P(A) \le P(B)$.

<u>*Доказательство*</u>. Так как $A \subset B$, то $A \cdot B=A$, поэтому равенство $B=A \cdot B+\overline{A} \cdot B$ дает равенство $B=A+\overline{A} \cdot B$. Поэтому $P(B)=P(A+\overline{A} \cdot B)=P(A)+P(\overline{A} \cdot B) \ge P(A)$.

□

<u>*Теоретические упражнения.*</u> 1) Доказать, что
$\boxed{P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1 \cdot A_2)-P(A_1 \cdot A_3)-P(A_2 \cdot A_3)+P(A_1 \cdot A_2 \cdot A_3)}$.

2) Вывести формулу для $P(A_1+A_2+...+A_n)$.

3) Доказать, что $P(A_1+A_2+...+A_n) \le P(A_1)+P(A_2)+...+P(A_n)$.

<u>*Пример 6*</u>. В книжном шкафу в случайном порядке расставлено 15 книг, причем 5 из них в твердом переплете. Студент берет наудачу 3 книги. Найти вероятность того, что хотя бы одна из взятых книг имеет твердый переплет.

<u>*Нерациональный, но поучительный способ решения.*</u> Пусть A – искомое событие (состоящее в том, что хотя бы одна из взятых книг имеет твердый переплет). Введем следующие события: A_1 – из взятых книг ровно одна имеет твердый переплет, A_2 – ровно две имеют твердый переплет, A_3 – ровно три. Ясно, что $A = A_1 + A_2 + A_3$, причем события A_1, A_2 и A_3 удовлетворяют условию, входящему в аксиому σ-аддитивности. Применяя эту аксиому, получаем:

$$P(A) = P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3).$$

Вероятности событий A_1, A_2 и A_3 находятся с помощью тех же рассуждений, которые были проведены при нахождении вероятности события A в примере 4. Таким образом получаем:

$$P\left(A_1\right) = \frac{C_5^1 \cdot C_{10}^2}{C_{15}^3} = \frac{45}{91}, \quad P\left(A_2\right) = \frac{C_5^2 \cdot C_{10}^1}{C_{15}^3} = \frac{20}{91}, \quad P\left(A_3\right) = \frac{C_5^3}{C_{15}^3} = \frac{2}{91}.$$

Окончательно, $P(A) = \frac{45}{91} + \frac{20}{91} + \frac{2}{91} = \frac{67}{91}$.

<u>*Рациональное решение (переход к противоположному событию).*</u> Событие A (состоящее в том, что хотя бы одна из взятых книг имеет твердый переплет) и событие $\overline{A}$ (состоящее в том, что ни одна из взятых книг не имеет твердого переплета) – противоположные события. По свойству 1 вероятности $P(A) = 1 - P(\overline{A})$. Вероятность события $\overline{A}$ вычисляется аналогично вероятности события A_3: $P(\overline{A}) = \frac{C_{10}^3}{C_{15}^3} = \frac{24}{91}$. Поэтому $P(A) = 1 - \frac{24}{91} = \frac{67}{91}$.

□

Независимые события и условные вероятности

При решении задач свойство 2 вероятности применяется чаще всего, если события A и B независимы. На интуитивном языке, событие A не зависит от события B, если реализация события A не зависит от того, совершилось событие B или нет. Математическая реализация интуитивного понятия независимости содержится в следующем определении.

<u>*Определение 6*</u>. События A и B на вероятностном пространстве $\left(\Omega, \mathcal{F}, P\right)$ называются независимыми, если

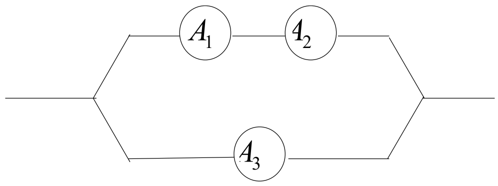
$$P(A \cdot B) = P(A) \cdot P(B). \tag{3}$$

События $A_1, A_2, \ldots, A_n, A_{n+1}, \ldots$ того же вероятностного пространства называются независимыми в совокупности (соответственно, попарно независимыми), если для любого конечного набора натуральных чисел $i_1 < i_2 < \ldots < i_k$ (соотв., для двух любых натуральных чисел $i_1 < i_2$) выполняется равенство $P\left(A_{i_1} \cdot A_{i_2} \cdot \ldots \cdot A_{i_k}\right) = P\left(A_{i_1}\right) \cdot P\left(A_{i_2}\right) \cdot \ldots \cdot P\left(A_{i_k}\right)$ (соотв., $P\left(A_{i_1} \cdot A_{i_2}\right) = P\left(A_{i_1}\right) \cdot P\left(A_{i_2}\right)$).

□

Определение 6 достаточно сложно для первого восприятия и мы к нему еще вернемся для более подробного обсуждения. А сейчас решим следующую задачу.

<u>*Пример 7*</u>. Рассмотрим часть электрической цепи:



Вероятность безотказной работы элемента A_1 равна 0,7, элемента A_2 – 0,8, элемента A_3 – 0,9. Найти вероятность того, что ток по цепи пройдет.

Обозначим через A_1, A_2 и A_3 события, состоящие в бесперебойной работе соответствующих элементов. Тогда $P(A_1)=0{,}7,\ P(A_2)=0{,}8,\ P(A_3)=0{,}9$. Пусть A – событие, состоящее в прохождении тока по цепи. Тогда $A=A_1\cdot A_2+A_3$. По свойству 2 вероятности имеем: $P(A)=P(A_1\cdot A_2+A_3)=P(A_1\cdot A_2)+P(A_3)-P(A_1\cdot A_2\cdot A_3)$. Так как элементы цепи работают независимо, то события A_1, A_2 и A_3 можно считать независимыми в совокупности в смысле определения 6. Поэтому $P(A_1\cdot A_2)=P(A_1)\cdot P(A_2)$ и $P(A_1\cdot A_2\cdot A_3)=P(A_1)\cdot P(A_2)\cdot P(A_3)$. Окончательно имеем:

$$P(A)=P(A_1)\cdot P(A_2)+P(A_3)-P(A_1)\cdot P(A_2)\cdot P(A_3)=0{,}56+0{,}9-0{,}504=0{,}956\,.$$ □

Следующий пример (разбор которого предоставляется читателю) показывает, что события могут быть попарно независимыми, но не быть независимыми в совокупности.

<u>*Пример 8*</u>. Пусть $\Omega=\{(x,y):\,0\le x\le 1, 0\le y\le 1\}$ – квадрат на плоскости xOy, $\mathcal{F}$ – σ-алгебра борелевских подмножеств этого квадрата (то есть наименьшая σ-алгебра, содержащая все прямоугольники, входящие в квадрат), а $P(A)=\textit{площадь}\,A$ для $A\in\mathcal{F}$. Рассмотрим события: $A_1=\left\{(x,y):\,x\le\frac{1}{2}\right\},\ A_2=\left\{(x,y):\,y\le\frac{1}{2}\right\},\ A_3=\left\{(x,y):\,\left(x-\frac{1}{2}\right)\cdot\left(y-\frac{1}{2}\right)\le 0\right\}$. Построить множества A_1, A_2, A_3. Доказать, что события A_1, A_2, A_3 попарно независимы, но не являются независимыми в совокупности.

□

<u>*Определение 7*</u>. Пусть A – фиксированное событие из σ-алгебры $\mathcal{F}$, такое, что $P(A)>0$. Для любого $B\in\mathcal{F}$ определим новую функцию:

$$P(B|A)=\frac{P(A\cdot B)}{P(A)}\,. \tag{4}$$

Эта функция называется условной вероятностью события B при условии, что событие A осуществилось (или просто *при условии* A).

□

Иногда целесообразно пользоваться следующим обозначением: $P^A(B)=P(B|A)$.

<u>*Теоретические упражнения.*</u> 1) Доказать, что для тройки $(\Omega,\mathcal{F},P^A)$ выполняются все аксиомы вероятностного пространства (см. определение 4). Эта тройка называется условным вероятностным пространством.

2)С помощью метода математической индукции доказать формулу:

$$\boxed{P(A_1\cdot A_2\cdot\ldots\cdot A_n)=P(A_1)\cdot P(A_2|A_1)\cdot\ldots\cdot P(A_n|A_1\cdot A_2\cdot\ldots\cdot A_{n-1})}.$$

Отметим следующие простые свойства независимости событий:

1. Достоверное событие Ω и невозможное событие $\varnothing$ независимы ни от какого другого события.

2. Несовместные события положительной вероятности зависимы.

3. Пусть $P(A)>0$. События A и B независимы тогда и только тогда, когда $P(B|A)=P(B)$.

ЧАСТЬ 1: ПЕРВОНАЧАЛЬНЫЕ ОПРЕДЕЛЕНИЯ И ФАКТЫ

Лекция 3

Пример 9. Урна содержит 3 красных и 2 черных шара (шары по форме неразличимы). Эксперимент состоит в том, что из урны наудачу последовательно вынимаются 2 шара. Пусть A – событие, состоящее в том, что первым изъят красный шар, а B – событие, состоящее в том, что вторым вынут красный шар. Найти условные вероятности $P(B|A)$ и $P(B|\overline{A})$.

Решение, использующее интуицию. Вычислим $P(B|A)$. Так как событие A совершилось, то первым мы вынули красный шар. Следовательно, в урне осталось 2 красных и 2 черных шара. После этого вероятность достать второй шар вычисляется по классической схеме: $P(B|A)=\frac{2}{4}=\frac{1}{2}$. Аналогично рассуждая, получаем $P(B|\overline{A})=\frac{3}{4}$.

Решение, использующее определение 7. Пометим красные шары цифрами 1, 2 и 3, а черные шары – цифрами 4 и 5. Очевидно, пространством элементарных событий Ω можно считать совокупность упорядоченных пар (ω_i,ω_j), где i,j=1,2,3,4,5 и $i\neq j$ (последнее условие проистекает из того, что если первым вынут шар ω_i, то так как в урну он не возвращается, то во второй раз его вытянуть нельзя). Так как в первый раз можно вытянуть один из пяти шаров, а второй раз – один из четырех, то $|\Omega|=5\cdot4=20$. Событию A благоприятствуют элементарные события (ω_i,ω_j), у которых i пробегает значения 1, 2 и 3, а j пробегает значения от 1 до 5, причем $i\neq j$. Очевидно, $|A|=3\cdot4=12$. Следовательно, $P(A)=\frac{|A|}{|\Omega|}=\frac{12}{20}=\frac{3}{5}$. Произведению $A\cdot B$ благоприятствуют пары (ω_i,ω_j), у которых i и j пробегают значения 1, 2 и 3 ($i\neq j$). Следовательно, $|A\cdot B|=3\cdot2=6$. Поэтому $P(A\cdot B)=\frac{|A\cdot B|}{|\Omega|}=\frac{6}{20}=\frac{3}{10}$. Применяя определение 7, получаем: $P(B|A)=\frac{P(A\cdot B)}{P(A)}=\frac{3}{10}:\frac{3}{5}=\frac{1}{2}$. Аналогично вычисляется $P(B|\overline{A})=0{,}75$.

□

Определение 8. Совокупность событий $\{H_1,H_2,...,H_n\}$ положительной вероятности на $(\Omega,\mathcal{F},P)$ называется полной группой гипотез, если:

1) события $\{H_1,H_2,...,H_n\}$ попарно несовместны (см. определение 5);

2) $H_1+H_2+...+H_n=\Omega$.

□

Теорема 1 (формула полной вероятности). Если $\{H_1,H_2,...,H_n\}$ – полная группа гипотез на $(\Omega,\mathcal{F},P)$, то вероятность любого события $A\in\mathcal{F}$ можно вычислить по формуле:

$$P(A)=P(A|H_1)\cdot P(H_1)+P(A|H_2)\cdot P(H_2)+...+P(A|H_n)\cdot P(H_n). \tag{5}$$

Для доказательства этой теоремы нам понадобится следующая лемма, представляющая и самостоятельный интерес.

Лемма 1. Если события $A_1,A_2,...,A_n,A_{n+1},...$ попарно несовместны (в смысле определения 5), то свойство σ-аддитивности из определения 4 выполняется, то есть

$$P(A_1+A_2+...+A_n+...)=P(A_1)+P(A_2)+...+P(A_n)+...$$

<u>*Доказательство леммы 1*</u>. В случае, если событий два, равенство $P(A_1+A_2)=P(A_1)+P(A_2)$ следует из свойства 2 вероятности. Если число событий конечно, то данное равенство можно доказать с помощью метода математической индукции. Если события образуют бесконечную последовательность, то доказательство несколько сложнее и предоставляется читателю в качестве необязательного теоретического упражнения.

□

<u>*Доказательство теоремы 1.*</u> Умножая обе части равенства $H_1+H_2+...+H_n=\Omega$ на событие *A*, получаем соотношение: $A\cdot\Omega=A\cdot H_1+A\cdot H_2+...+A\cdot H_n$. Ясно, что $A\cdot\Omega=A$. Нетрудно видеть также, что события $A\cdot H_1, A\cdot H_2, ..., A\cdot H_n$ попарно несовместны. Докажем это, например, для событий $A\cdot H_1$ и $A\cdot H_2$. Имеем:

$$P\big((A\cdot H_1)\cdot(A\cdot H_2)\big)=P\big(A\cdot A\cdot H_1\cdot H_2\big)=P\big(A\cdot H_1\cdot H_2\big)\le P\big(H_1\cdot H_2\big)=0$$

(в этой цепочке неравенство мы записали в силу свойства 3 вероятности). Итак, $P\big((A\cdot H_1)\cdot(A\cdot H_2)\big)\le 0$. Так как всегда $P\big((A\cdot H_1)\cdot(A\cdot H_2)\big)\ge 0$, то $P\big((A\cdot H_1)\cdot(A\cdot H_2)\big)=0$, что и доказывает несовместность событий $A\cdot H_1$ и $A\cdot H_2$.

Применяя лемму 1, получаем: $P(A)=P(A\cdot\Omega)=P(A\cdot H_1)+P(A\cdot H_2)+...+P(A\cdot H_n)$. Принимая во внимание формулу (4), имеем:

$$P(A\cdot H_1)=P(A|H_1)\cdot P(H_1),\ P(A\cdot H_2)=P(A|H_2)\cdot P(H_2),...,P(A\cdot H_n)=P(A|H_n)\cdot P(H_n),$$

откуда и вытекает формула (5). □

<u>*Пример 10*</u>. Пусть мы находимся в рамках условий примера 9. Требуется найти $P(B)$.

Заметим прежде всего, что в отличие от примера 9 нам ничего не известно о том, совершилось ли событие *A* или нет. Поэтому естественно ввести гипотезы $H_1=A$ и $H_2=\overline{A}$. Очевидно, что $\{H_1, H_2\}$ – полная группа гипотез. По формуле (5) полной вероятности имеем:

$$P(B)=P\big(B|H_1\big)\cdot P\big(H_1\big)+P\big(B|H_2\big)\cdot P\big(H_2\big).$$

Присутствующие здесь условные вероятности мы уже вычислили в примере 9, а именно: $P\big(B|H_1\big)=P\big(B|A\big)=1/2$, $P\big(B|H_2\big)=P\big(B|\overline{A}\big)=3/4$. Используя классическое определение вероятности, имеем: $P\big(H_1\big)=P\big(A\big)=3/5$, $P\big(H_2\big)=P\big(\overline{A}\big)=2/5$. Следовательно, $P(B)=\frac{1}{2}\cdot\frac{3}{5}+\frac{3}{4}\cdot\frac{2}{5}=0{,}6$. □

<u>*Пример 11*</u>. По цели произведено три выстрела. Вероятность попасть в цель при первом выстреле (событие A_1) равна $p_1=0{,}3$, при втором выстреле (событие A_2) – $p_2=0{,}6$, при третьем выстреле (событие A_3) – $p_3=0{,}8$. Вероятность разрушения цели при одном попадании равна $0{,}4$, при двух попаданиях – 0,7, при трех попаданиях – 1. Определить вероятность разрушения цели после произведения трех выстрелов (событие *A*).

Введем систему гипотез: H_1 – произошло одно попадание в цель, H_2 – произошло два попадания в цель, H_3 – три попадания, H_4 – нет попаданий. Ясно, что $\{H_1, H_2, H_3, H_4\}$ есть полная группа гипотез. Вычислим вероятности этих гипотез.

В цель осуществлено ровно одно попадание, если при первом выстреле было попадание, а при втором и третьем – промах; или при втором выстреле было попадание, а при первом и третьем – промах; или при третьем выстреле было попадание, а при первом и втором – промах. Поэтому $H_1=A_1\cdot\overline{A_2}\cdot\overline{A_3}+\overline{A_1}\cdot A_2\cdot\overline{A_3}+\overline{A_1}\cdot\overline{A_2}\cdot A_3$, причем ясно, что слагаемы в данной сумме попарно несовместны. Следовательно,

$$P\big(H_1\big)=P\big(A_1\cdot\overline{A_2}\cdot\overline{A_3}\big)+P\big(\overline{A_1}\cdot A_2\cdot\overline{A_3}\big)+P\big(\overline{A_1}\cdot\overline{A_2}\cdot A_3\big).$$

Так как первый, второй и третий выстрелы никак не связаны друг с другом, то можно считать, что события A_1, A_2 и A_3 независимы в совокупности. Из этого следует, что независимы в совокупности и тройки событий $\left\{A_1, \overline{A_2}, \overline{A_3}\right\}, \left\{\overline{A_1}, A_2, \overline{A_3}\right\}, \left\{\overline{A_1}, \overline{A_2}, A_3\right\}$ (докажите это!). То есть

$$P(H_1)=P(A_1)\cdot P(\overline{A_2})\cdot P(\overline{A_3})+P(\overline{A_1})\cdot P(A_2)\cdot P(\overline{A_3})+P(\overline{A_1})\cdot P(\overline{A_2})\cdot P(A_3).$$

Следовательно,

$$P(H_1)=p_1(1-p_2)(1-p_3)+(1-p_1)p_2(1-p_3)+(1-p_1)(1-p_2)p_3=0{,}332.$$

Рассуждая аналогично, получаем:

$$P(H_2)=p_1p_2(1-p_3)+p_1(1-p_2)p_3+(1-p_1)p_2p_3=0{,}468,$$

$$P(H_3)=p_1p_2p_3=0{,}144,$$

$$P(H_4)=(1-p_1)(1-p_2)(1-p_3)=0{,}056.$$

Запишем теперь условные вероятности разрушения цели при выполнении соответствующих гипотез. Эти вероятности заданы в условии:

$$P(A|H_1)=0{,}4,\ P(A|H_2)=0{,}7,\ P(A|H_3)=1,\ P(A|H_4)=0.$$

Применяя формулу (5) полной вероятности, получаем:

$$P(A)=P(A|H_1)\cdot P(H_1)+P(A|H_2)\cdot P(H_2)+P(A|H_3)\cdot P(H_3)+P(A|H_4)\cdot P(H_4)=$$
$$=0{,}4\cdot 0{,}332+0{,}7\cdot 0{,}468+1\cdot 0{,}144+0\cdot 0{,}056=0{,}6044.$$

□

Представим теперь себе ситуацию, что событие *A* осуществилось. Это обстоятельство позволяет пересмотреть вероятности исходных гипотез $\{H_1, H_2, ..., H_n\}$, то есть вычислить так называемые *апостериорные* вероятности гипотез (в отличие от *априорных* вероятностей, которые получаются до реализации какого-либо события). Речь идет об условных вероятностях $P(H_1|A), P(H_2|A), ..., P(H_n|A)$.

<u>*Теорема 2*</u> (формула Байеса). Для любого $k=1,2,...,n$ справедлива формула:

$$P(H_k|A)=\frac{P(A|H_k)\cdot P(H_k)}{\sum_{i=1}^{n}P(A|H_i)\cdot P(H_i)}. \tag{6}$$

<u>*Доказательство*</u>. Применим формулу (4) в двух вариантах: $P(A\cdot H_k)=P(H_k|A)\cdot P(A)$ и $P(A\cdot H_k)=P(A|H_k)\cdot P(H_k)$. Поэтому

$$P(H_k|A)\cdot P(A)=P(A|H_k)\cdot P(H_k)\ \Rightarrow\ P(H_k|A)=\frac{P(A|H_k)\cdot P(H_k)}{P(A)}.$$

Записывая для $P(A)$ формулу (5) полной вероятности, получаем соотношение (6).

□

<u>*Пример 12*</u> (продолжение примера 10). Пусть в условиях примеров 9 и 10 осуществилось событие *B* (то есть во второй раз из урны вынули красный шар). Какова вероятность того, что первым был вынут также красный шар?

Так как $A=H_1$, то нам нужно вычислить условную вероятность $P(H_1|B)$. По формуле Байеса (6) имеем:

$$P(H_1|B)=\frac{P(B|H_1)\cdot P(H_1)}{P(B)}=\frac{0{,}5\cdot 0{,}6}{0{,}6}=0{,}5.$$

□

В рассмотренном примере условные вероятности $P(H_1|B)$ и $P(B|H_1)$ оказались равными. Это получилось совершенно случайно и не отражает никакой закономерности.

<u>*Пример13*</u> (продолжение примера 11). Пусть в условиях примера 11 цель разрушена (то есть совершилось событие *A*). Какова вероятность того, что при этом в цель было два попадания?

Очевидно, наша задача – вычислить условную вероятность $P(H_2|A)$. Имеем по формуле (6):

$$P(H_2|A)=\frac{P(A|H_2)\cdot P(H_2)}{P(A)}=\frac{0{,}7\cdot 0{,}468}{0{,}6044}\approx 0{,}542.$$

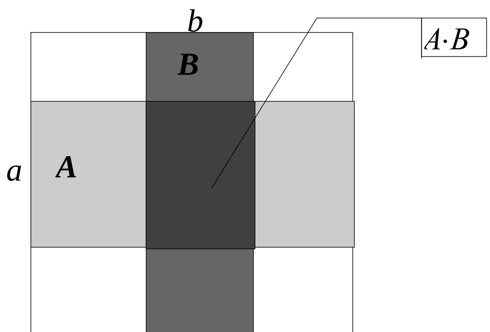
□

Геометрическая вероятность

Рассмотрим теперь более подробно вероятностное пространство $(\Omega,\mathcal{F},P)$, введенное в примере 8. Лучше всего с этим пространством связывать эксперимент стрельбы в квадрат Ω (считается, что в этот квадрат стрелок попадает при любых условиях). Любое борелевское (см. пример 8) подмножество $A\in\mathcal{F}$ мы в этом случае интерпретируем как событие, состоящее в том, что стрелок попал в *A*. Смысл рассматриваемой вероятности *P* (или, как часто говорят, геометрической вероятности) хорошо согласуется с интуицией: чем больше площадь подмножества *A*, тем больше вероятность наступления события *A*.

Используя данное вероятностное пространство, можно достаточно легко получать геометрические обоснования многих фактов, доказанных нами ранее аналитически. Например, свойство 1 вероятности геометрически очевидно: площадь дополнительного к *A* множества $\overline{A}$ равна площади квадрата Ω минус площадь самого множества *A*. Свойство 2 вероятности также становится прозрачным: так как $P(A+B)$ это заштрихованная площадь на первом рисунке (см. пункт "Действия над событиями. Алгебра событий"), а в выражение $P(A)+P(B)$ два раза входит площадь множества $A\cdot B$, то ее нужно один раз отнять, и в результате получается формула $P(A+B)=P(A)+P(B)-P(A\cdot B)$. Точно так же легко интерпретируется свойство 3 вероятности, так как фраза "из события *A* следует событие *B*" геометрически означает, что множество *A* содержится во множестве *B*.

Проинтерпретируем теперь понятие независимости событий (см. определение 6). Пусть события *A* и *B* таковы, какими они показаны на следующем рисунке:



Здесь множество *A* представляет собой прямоугольник с шириной, равной 1, и высотой, равной *a*. Множество *B* представляет собой прямоугольник с шириной *b* и высотой, равной 1. Множество $A\cdot B$ – это пересечение прямоугольников *A* и *B* и также является прямоугольником. Заметим, что стороны всех полученных прямоугольников параллельны соответствующим сторонам квадрата Ω. Из геометрических соображений следует, что $P(A)=1\cdot a=a,\ P(B)=b\cdot 1=b,\ P(A\cdot B)=a\cdot b$. Поэтому выполняется условие независимости событий $P(A\cdot B)=P(A)\cdot P(B)$, то есть события *A* и *B* независимы. Оказывается, события такой конфигурации, как *A* и *B*, дают самые характерные примеры независимых событий. Подобная процедура построения независимых событий может быть существенно обобщена. Эти обобщения сплошь и рядом используются в теории вероятностей.

ЧАСТЬ 2: СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Лекция 4

Случайные величины на конечных вероятностных пространствах

<u>*Определение 9*</u>. ***Конечным вероятностным пространством*** называется такое вероятностное пространство $(\Omega, \mathcal{F}, P)$ (см. определение 4), у которого пространство элементарных событий Ω состоит из конечного числа элементов, а σ-алгебра $\mathcal{F}$ представляет собой совокупность ***всех*** подмножеств множества Ω.

□

Большинство примеров, рассмотренных в первой части наших лекций, было реализовано на конечных вероятностных пространствах. Заметим, что классическое вероятностное пространство (см. определение 1) есть частный случай конечного вероятностного пространства.

<u>*Определение 10*</u>. Пусть дано конечное вероятностное пространство $(\Omega, \mathcal{F}, P)$. Числовая функция $X = X(\omega)$, определенная на пространстве элементарных событий Ω, называется случайной величиной (с.в.) на $(\Omega, \mathcal{F}, P)$.

□

Для обозначения случайных величин мы, в основном, применяем большие латинские буквы X, Y, Z и т.д., но иногда будем использовать и малые греческие буквы.

<u>*Пример 14*</u> (продолжение примера 1). Пусть мы находимся в условиях примера 1. Здесь мы имеем классическое вероятностное пространство, у которого $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, где ω_k – элементарное событие, заключающееся в выпадении числа *k* (*k*=1,2,...,6). Тогда функция $X = X(\omega)$, заданная формулой $X(\omega_k) = k$, (*k*=1,2,...,6), является с.в. на данном вероятностном пространстве.

□

<u>*Определение 11*</u>. Пусть *A* – событие из σ-алгебры $\mathcal{F}$. Индикатором события *A* называется случайная величина I_A, заданная формулой:

$$I_A(\omega) = \begin{cases} 1, \text{ если } \omega \in A, \\ 0, \text{ если } \omega \notin A. \end{cases} \tag{7}$$

□

Перечислим некоторые легко доказываемые свойства индикаторов:

1) $I_\varnothing \equiv 0,\ I_\Omega \equiv 1$;

2) $I_{A \cdot B} = I_A \cdot I_B$;

3) $I_{\overline{A}} = 1 - I_A$;

4) если $A \cdot B = \varnothing$, то $I_{A+B} = I_A + I_B$.

Заметим, что определение 11 и записанные свойства индикаторов имеют силу на произвольных вероятностных пространствах (а не только конечных).

<u>*Лемма 2*</u> (о представлении с.в. через индикаторы). Пусть $X = X(\omega)$ – с.в. на конечном вероятностном пространстве $(\Omega, \mathcal{F}, P)$ и $x_1 < x_2 < ... < x_m$ – всевозможные значения этой с.в. Пусть событие H_k состоит из всех $\omega \in \Omega$, для которых $X(\omega) = x_k$, $k = 1,2,...,m$, то есть $H_k = \{\omega \in \Omega : X(\omega) = x_k\}$. Тогда:

$$X(\omega) = \sum_{k=1}^{m} x_k \cdot I_{H_k}. \tag{8}$$

Доказательство непосредственно вытекает из того факта, что система событий $\{H_1, H_2, .., H_m\}$ образует полную группу гипотез (см. определение 8). Детали доказательства предлагается провести читателю.

□

Схема Бернулли

Предположим, что мы производим некоторое испытание с двумя исходами (например, бросание монеты, когда исходами являются орел или решка, или вытягивание лотерейного билета, когда в результате испытания билет оказывается выигрышным либо проигрышным). Один исход данного испытания (имеющий вероятность p) будем считать успехом и обозначать единицей. Второй исход (имеющий вероятность $q=1-p$) будем считать неудачей и обозначать нулем. Таким образом, совокупность исходов данного испытания мы отождествили с двуточечным множеством $\{0;1\}$.

Повторим теперь это испытание независимым образом n раз. Результатом этого n-кратного эксперимента будут последовательности вида $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, где каждое число ε_k ($k=1,2,...,n$) равно нулю либо единице. Например, если $n=4$ и мы получили последовательность $(0,1,1,0)$, то это означает, что при первом и четвертом испытаниях нас постигла неудача, а второе и третье испытания были успешными. Ясно, что в качестве пространства элементарных событий следует взять множество Ω, состоящее из ***всевозможных*** цепочек вида $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, где каждое число ε_k ($k=1,2,...,n$) равно нулю либо единице. В качестве σ-алгебры F выберем совокупность всех подмножеств множества Ω.

Остановимся более подробно на определении вероятности P. Обозначим через A_k^0 (соответственно A_k^1) событие, состоящее в том, что при k-м испытании нас постигает неудача (соотв., при k-м испытании мы имеем успех). Очевидно, $A_k^0 = \{(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) \colon \varepsilon_k = 0\}$, $A_k^1 = \{(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) \colon \varepsilon_k = 1\}$. Так как при разных k испытания производятся независимым образом, то математически это должно означать, что любая система событий $\{A_1^{\varepsilon_1}, A_2^{\varepsilon_2}, ..., A_n^{\varepsilon_n}\}$ должна быть независимой в совокупности в смысле определения 6. То есть вероятность события $A_1^{\varepsilon_1} \cdot A_2^{\varepsilon_2} \cdot ... \cdot A_n^{\varepsilon_n}$ ***должна*** определяться формулой:

$$P\left(A_1^{\varepsilon_1} \cdot A_2^{\varepsilon_2} \cdot ... \cdot A_n^{\varepsilon_n}\right) = P\left(A_1^{\varepsilon_1}\right) \cdot P\left(A_2^{\varepsilon_2}\right) \cdot ... \cdot P\left(A_n^{\varepsilon_n}\right) = p^{\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n} \cdot q^{n - (\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n)}. \quad (9)$$

Но легко видеть, что событие $A_1^{\varepsilon_1} \cdot A_2^{\varepsilon_2} \cdot ... \cdot A_n^{\varepsilon_n}$ совпадает с элементарным событием $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, поэтому вероятность каждого элементарного события $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ должна определяться той же формулой:

$$P\{(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)\} = p^{\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n} \cdot q^{n - (\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n)}. \quad (10)$$

Таким образом, мы построили конечное вероятностное пространство (Ω, F, P) (бернуллиевское вероятностное пространство), моделирующее описанный выше n-кратный эксперимент. Определим теперь на (Ω, F, P) случайную величину μ_n, равную числу успехов при n независимых испытаниях в рассмотренной нами "схеме Бернулли", то есть если $\omega = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, то

$$\mu_n(\omega) = \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n. \quad (11)$$

Очевидно, μ_n может принимать значения $0,1,2,...,n$. Обозначим $H_m = \{\omega \in \Omega \colon \mu_n(\omega) = m\}$. Событие H_m состоит в том, что в n испытаниях в схеме Бернулли наступило ровно m успехов.

<u>Теорема 3</u>. Справедлива формула:

$$P(H_m)=C_n^m \cdot p^m q^{n-m}, \tag{12}$$

где C_n^m – число сочетаний из n элементов по m (см. формулу (2)).

<u>Доказательство</u>. Из определения события H_m и из свойства аддитивности (см. определение 4) следует, что $P(H_m)=\sum\limits_{\omega\in H_m} P\{\omega\}$. Ввиду соотношения (10), все вероятности, находящиеся под знаком суммирования, одинаковы и равны $p^m q^{n-m}$. Поэтому $P(H_m)=|H_m|\cdot p^m q^{n-m}$, где $|H_m|$ – число элементов множества H_m. Число $|H_m|$, очевидно, совпадает с количеством способов выбора m мест для "1" в цепочке $\omega=(\varepsilon_1,\varepsilon_2,...,\varepsilon_n)$ при том предположении, что оставшиеся места автоматически заполняются нулями. Следовательно, $|H_m|=C_n^m$ и формула (12) доказана.

□

<u>Пример 15</u>. Игральная кость бросается десять раз. Найти вероятность того, что при этом выпало ровно три шестерки.

При каждом бросании будем считать успехом выпадение шестерки, а неудачей – выпадение любого другого числа. Тогда мы будем находиться в рамках схемы Бернулли с $n=10, m=3, p=\frac{1}{6}, q=\frac{5}{6}$. Используя формулу (12), получаем:

$$P(H_3)=C_{10}^3\cdot\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7=\frac{10\cdot 9\cdot 8\cdot 5^7}{1\cdot 2\cdot 3\cdot 6^{10}}\approx 0{,}155.$$

□

Закон распределения случайной величины в случае конечного вероятностного пространства

<u>Определение 12</u>. Пусть $X=X(\omega)$ – с.в. на конечном вероятностном пространстве $(\Omega,\mathcal{F},P)$ и $x_1<x_2<...<x_m$ – всевозможные значения этой с.в. ***Законом распределения*** данной с.в. называется таблица

x_1	x_2	...	x_m
p_1	p_2	...	p_m

,

где нижний ряд состоит из чисел p_k, равных вероятностям принятия случайной величиной $X=X(\omega)$ значений x_k, то есть $p_k=P(\omega\in\Omega: X(\omega)=x_k)$, причем $p_1+p_2+...+p_m=1$.

□

<u>Пример 16</u> (продолжение примера 14). Очевидно, закон распределения случайной величины, определенной в примере 14, имеет вид:

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

.

□

<u>Пример 17</u>. Закон распределения индикатора I_A имеет вид:

I_A	0	1
P	$1-P(A)$	$P(A)$

.

□

<u>Пример 18</u>. С.в. μ_n, равная числу успехов в серии из n независимых испытаний, в силу формулы (12) имеет следующий закон распределения:

μ_n	0	1	2	...	n
P	q^n	npq^{n-1}	$\frac{n(n-1)}{2}\cdot p^2q^{n-2}$	...	p^n

.

Этот закон распределения называется биномиальным.

□

Математическое ожидание случайной величины
в случае конечного вероятностного пространства

<u>*Определение 13*</u>. Математическим ожиданием с.в. $X=X(\omega)$, определенной на конечном вероятностном пространстве $(\Omega,\mathcal{F},P)$, называется число

$$MX=\sum_{\omega\in\Omega}X(\omega)\cdot P\{\omega\}\,, \qquad (13)$$

где при суммировании ω пробегает все пространство элементарных событий Ω, а $P\{\omega\}$ есть вероятность элементарного события ω.

□

Если перенумеровать все элементарные события из Ω, то есть записать $\Omega=\{\omega_1,\omega_2,\ldots,\omega_n\}$, то математическое ожидание запишется в более прозрачном виде:

$$MX=\sum_{k=1}^{n}X(\omega_k)\cdot P\{\omega_k\}=X(\omega_1)\cdot P\{\omega_1\}+X(\omega_2)\cdot P\{\omega_2\}+\ldots+X(\omega_n)\cdot P\{\omega_n\}\,.$$

Однако для выкладок формула (13) предпочтительнее.

Свойства математического ожидания

1. Математическое ожидание обладает свойством линейности:

$$M(\alpha\cdot X+\beta\cdot Y)=\alpha\cdot MX+\beta\cdot MY\,.$$

<u>*Доказательство*</u>. Применяя определение 13, получаем:

$$M(\alpha\cdot X+\beta\cdot Y)=\sum_{\omega\in\Omega}(\alpha\cdot X(\omega)+\beta\cdot Y(\omega))\cdot P\{\omega\}=\alpha\cdot\sum_{\omega\in\Omega}X(\omega)\cdot P\{\omega\}+\beta\cdot\sum_{\omega\in\Omega}Y(\omega)\cdot P\{\omega\}=\alpha\cdot MX+\beta\cdot MY\,.$$ □

2. Математическое ожидание индикатора события A равно вероятности этого события:

$$MI_A=P(A)\,.$$

<u>*Доказательство*</u>. Так как $P(A)=\sum_{\omega\in A}P\{\omega\}$, то

$$MI_A=\sum_{\omega\in\Omega}I_A(\omega)\cdot P\{\omega\}=\sum_{\omega\in A}I_A(\omega)\cdot P\{\omega\}+\sum_{\omega\in\bar{A}}I_A(\omega)\cdot P\{\omega\}=\sum_{\omega\in A}1\cdot P\{\omega\}+\sum_{\omega\in\bar{A}}0\cdot P\{\omega\}=\sum_{\omega\in A}P\{\omega\}=P(A)\,.$$

□

3. Математическое ожидание обладает свойством монотонности: если $X\geq Y$, то $MX\geq MY$.

<u>*Доказательство*</u> провести самостоятельно.

□

4. Если с.в. X обладает законом распределения, выписанным в определении 12, то справедлива формула:

$$MX=\sum_{k=1}^{m}x_k\cdot p_k\,, \qquad (14)$$

<u>*Доказательство*</u>. Применив к формуле (8) операцию математического ожидания и используя свойство линейности, а также формулу для вычисления математического ожидания от индикатора, получим:

$$MX=M\left(\sum_{k=1}^{m}x_k\cdot I_{H_k}\right)=\sum_{k=1}^{m}x_k\cdot MI_{H_k}=\sum_{k=1}^{m}x_k\cdot P(H_k)=\sum_{k=1}^{m}x_k\cdot p_k\,.$$

□

ЧАСТЬ 2: СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Лекция 5

Вычисление математических ожиданий

Покажем на примере, как для вычисления математического ожидания используется формула (14).

<u>*Пример19*</u> (продолжение примера 16). Используя закон распределения с.в. *X*, равной числу очков, выпавших на игральной кости, получаем:

$$MX=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{1}{6}\cdot(1+2+3+4+5+6)=\frac{1}{6}\cdot\frac{1+6}{2}\cdot 6=3{,}5\,.$$

□

Иногда при вычислении математических ожиданий случайную величину представляют в виде суммы более простых случайных величин, а затем применяют свойство линейности.

<u>*Пример 20*</u> (продолжение примера 18). Используя закон распределения с.в. μ_n, матожидание $M\mu_n$ можно было бы опять вычислить по формуле (14) (проделайте это!). Однако значительно проще поступить следующим образом. Рассмотрим событие A_k^1, состоящее в том, что в *k*-м испытании произошел успех *k*=1,2,...,*m* (см. раздел "Схема Бернулли"). Очевидно, справедливо представление:

$$\mu_n=I_{A_1^1}+I_{A_2^1}+...+I_{A_n^1}\,. \qquad (15)$$

По свойству линейности имеем:

$$M\mu_n=MI_{A_1^1}+MI_{A_2^1}+...+MI_{A_n^1}\,.$$

Применяя теперь свойство 2 математического ожидания, получим:

$$M\mu_n=P\left(A_1^1\right)+P\left(A_2^1\right)+...+P\left(A_n^1\right)=p+p+...+p=np\,.$$

□

Следующие теоремы, доказательства которых мы опускаем, позволяют вычислять математические ожидания функций от случайных величин.

<u>*Теорема 4*</u>. Пусть *X* – с.в. с законом распределения, выписанным в определении 12, а $f(x)$ – некоторая функция действительного переменного. Тогда:

$$Mf(X)=\sum_{k=1}^{m} f(x_k)\cdot p_k\,. \qquad (16)$$

□

<u>*Пример 21*</u> (продолжение примера 19). Пусть $f(x)=x^2$. Тогда формула (16) дает:

$$M\left(X^2\right)=1^2\cdot\frac{1}{6}+2^2\cdot\frac{1}{6}+3^2\cdot\frac{1}{6}+4^2\cdot\frac{1}{6}+5^2\cdot\frac{1}{6}+6^2\cdot\frac{1}{6}=\frac{1}{6}\cdot\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\frac{91}{6}\,.$$

□

<u>*Теорема 5*</u>. Пусть *X* и *Y* – случайные величины на конечном вероятностном пространстве $(\Omega,\mathcal{F},P)$, принимающие соответственно значения $x_1<x_2<...<x_m$ и $y_1<y_2<...<y_n$, а $g(x,y)$ – некоторая функция двух действительных переменных. Тогда

$$Mg(X,Y)=\sum_{k=1}^{m}\sum_{i=1}^{n} g\left(x_k,y_i\right)\cdot P\left(X=x_k,\,Y=y_i\right). \qquad (17)$$

□

Заметим, что для вычисления $Mg(X,Y)$, вообще говоря, недостаточно знать законы распределения случайных величин *X* и *Y*, но нужно иметь таблицу чисел $P\left(X=x_k,\,Y=y_i\right)$, которую называют ***совместным*** распределением с.в. X и Y. Однако, когда с.в. X и Y ***независимы***, ситуация упрощается.

Определение 14. С.в. X и Y (см. формулировку теоремы 5) называются независимыми, если для любых k=1,2,...,m и i=1,2,...,n выполняются равенства:

$$P\left(X=x_k,\, Y=y_i\right)=P\left(X=x_k\right)\cdot P\left(Y=y_i\right). \tag{18}$$

□

Теорема 6. Математическое ожидание произведения независимых с.в. равно произведению их математических ожиданий, то есть если с.в. X и Y независимы, то

$$M(X\cdot Y)=M(X)\cdot M(Y). \tag{19}$$

Доказательство. Пусть $g(x,y)=x\cdot y$. Тогда формула (17) примет вид:

$$M(X\cdot Y)=\sum_{k=1}^{m}\sum_{i=1}^{n}x_k\,y_i\cdot P\left(X=x_k,\, Y=y_i\right).$$

Так как с.в. X и Y независимы, то можно применить формулу (18) и представить двойную сумму как произведение одинарных сумм:

$$M(X\cdot Y)=\sum_{k=1}^{m}\sum_{i=1}^{n}x_k\,y_i\cdot P\left(X=x_k\right)P\left(Y=y_i\right)=\left(\sum_{k=1}^{m}x_k\cdot P\left(X=x_k\right)\right)\cdot\left(\sum_{i=1}^{n}y_i\cdot P\left(Y=y_i\right)\right)=MX\cdot MY.$$

□

Лемма 3. Если события A и B независимы, то индикаторы I_A и I_B этих событий есть независимые с.в.

Доказательство. Прежде всего докажем, что из независимости событий A и B следует независимость событий A и $\overline{B}$ (этот факт мы уже использовали в примере 11). Действительно, так как $A=A\cdot B+A\cdot\overline{B}$, причем слагаемые в этом равенстве несовместны, то $P(A)=P(A\cdot B)+P(A\cdot\overline{B})$. По определению 6 независимости событий получаем $P(A\cdot B)=P(A)\cdot P(B)$, то есть $P(A)=P(A)P(B)+P(A\cdot\overline{B})$. Отсюда без труда выводим: $P(A\cdot\overline{B})=P(A)-P(A)P(B)=P(A)\cdot(1-P(B))=P(A)\cdot P(\overline{B})$, то есть события A и $\overline{B}$ независимы. Аналогично доказывается независимость событий $\overline{A}$ и B, а также $\overline{A}$ и $\overline{B}$. Поэтому:

$$P\left(I_A=1, I_B=1\right)=P(A\cdot B)=P(A)\cdot P(B)=P\left(I_A=1\right)\cdot P\left(I_B=1\right),$$
$$P\left(I_A=1, I_B=0\right)=P\left(A\cdot\overline{B}\right)=P(A)\cdot P\left(\overline{B}\right)=P\left(I_A=1\right)\cdot P\left(I_B=0\right),$$
$$P\left(I_A=0, I_B=1\right)=P\left(\overline{A}\cdot B\right)=P\left(\overline{A}\right)\cdot P(B)=P\left(I_A=0\right)\cdot P\left(I_B=1\right),$$
$$P\left(I_A=0, I_B=0\right)=P\left(\overline{A}\cdot\overline{B}\right)=P\left(\overline{A}\right)\cdot P\left(\overline{B}\right)=P\left(I_A=0\right)\cdot P\left(I_B=0\right).$$

Из полученных равенств следует независимость индикаторов I_A и I_B (см. определение 14).

□

Пример 22. Заимствуя обозначения из примера 20, рассмотрим событие $A_{k-1}^1\cdot A_k^1$, состоящее в том, что на k-м испытании произошел двойной успех. С.в. $\nu_n=\sum_{k=2}^{n}I_{A_{k-1}^1\cdot A_k^1}$ равна числу двойных успехов в серии из n испытаний в схеме Бернулли. Вычислим $M\nu_n$. Имеем:

$$M\nu_n=\sum_{k=2}^{n}M\left(I_{A_{k-1}^1\cdot A_k^1}\right)=\sum_{k=2}^{n}M\left(I_{A_{k-1}^1}\cdot I_{A_k^1}\right).$$

Так как события A_{k-1}^1 и A_k^1 независимы, то по лемме 3 индикаторы $I_{A_{k-1}^1}$ и $I_{A_k^1}$ также независимы. Применяя теорему 6, получаем:

$$M\nu_n=\sum_{k=2}^{n}M\left(I_{A_{k-1}^1}\right)\cdot M\left(I_{A_k^1}\right)=\sum_{k=2}^{n}P\left(A_{k-1}^1\right)\cdot P\left(A_k^1\right)=\sum_{k=2}^{n}p^2=(n-1)p^2.$$

□

Дисперсия случайной величины в случае конечного вероятностного пространства

<u>*Определение 15*</u>. Дисперсией с.в. $X=X(\omega)$ называется число

$$DX=M\left[(X-MX)^2\right]. \tag{20}$$

□

Таким образом, дисперсия – это математическое ожидание квадрата отклонения с.в. *X* от ее математического ожидания.

Свойства дисперсии

1. Справедлива следующая формула для вычисления дисперсии:

$$DX=M\left(X^2\right)-(MX)^2. \tag{21}$$

<u>*Доказательство*</u>. Временно обозначим $MX=C$. Тогда, применяя свойство линейности математического ожидания, получаем:

$$DX=M\left[(X-MX)^2\right]=M\left[(X-C)^2\right]=M\left[X^2-2CX+C^2\right]=M\left(X^2\right)-2C\cdot MX+C^2.$$

Заменяя *C* через MX, получаем формулу (21).

□

2. Для любой с.в. *X* справедливо неравенство $DX\geq 0$. Это неравенство превращается в равенство тогда и только тогда, когда с.в. *X* постоянна с вероятностью единица.

<u>*Доказательство*</u>. Так как $(X-MX)^2\geq 0$, то по свойству 3 математического ожидания (свойство монотонности) $DX=M\left[(X-MX)^2\right]\geq M[0]=0$.

Если с.в. постоянна (то есть $X\equiv C$), то $MX=C$, и следовательно, $DX=M[0]=0$. Обратно, пусть $DX=0$. Обозначим $MX=C$ и предположим, что с.в. *X* имеет закон распределения, выписанный в определении 12, причем все $p_k>0$ и $\sum_{k=1}^{m}p_k=1$. Рассмотрим функцию $f(x)=(x-C)^2$. Применяя формулу (16), получаем:

$$0=DX=M\left[(X-MX)^2\right]=M[f(X)]=\sum_{k=1}^{m}f(x_k)\cdot p_k. \tag{22}$$

Ясно, что $f(x_k)=(x_k-C)^2\geq 0$ при всех $k=1,2,...,m$, то есть все члены последней суммы неотрицательны. Если хотя бы для одного значения *k* выполнялось строгое неравенство, то мы получили бы, что $\sum_{k=1}^{m}f(x_k)\cdot p_k>0$, в противоречии с равенством (22). Поэтому при всех $k=1,2,...,m$ $f(x_k)=0$, то есть $(x_k-C)^2=0$, а следовательно, $x_k=C$. Но так как $\sum_{k=1}^{m}p_k=1$, то это как раз и означает, что $X=C$ с вероятностью 1.

□

3. Пусть *c* – произвольное действительное число. Тогда

$$D(cX)=c^2\cdot DX. \tag{23}$$

<u>*Доказательство*</u>. Применяя формулу (21), получаем:

$$D(cX)=M\left[(cX)^2\right]-\left[M(cX)\right]^2=M\left[c^2\cdot X^2\right]-\left[c\cdot MX\right]^2=c^2\cdot M\left(X^2\right)-c^2\cdot(MX)^2=c^2\cdot\left[M\left(X^2\right)-(MX)^2\right]=$$
$$=c^2\cdot DX.$$

□

4. Если с.в. *X* и *Y* независимы (см. определение 14), то

$$D(X+Y)=DX+DY. \tag{24}$$

Доказательство этого факта основано на следующей лемме, обоснование которой предоставляется читателю.

Лемма 4. Если с.в. X и Y независимы, а C_1 и C_2 – действительные числа, то с.в. $X+C_1$ и $Y+C_2$ также независимы.

□

Доказательство свойства 4. Используя определение 15, а также свойство линейности математического ожидания, получаем:

$$D(X+Y)=M\left[(X+Y)-M(X+Y)\right]^2=M\left[(X-MX)+(Y-MY)\right]^2=$$
$$=M\left[(X-MX)^2+2\cdot(X-MX)\cdot(Y-MY)+(Y-MY)^2\right]=$$
$$=M(X-MX)^2+2\cdot M\left[(X-MX)\cdot(Y-MY)\right]+M(Y-MY)^2.$$

Так как MX и MY – постоянные действительные числа, то из леммы 4 следует, что с.в. $X-MX$ и $Y-MY$ независимы. Поэтому по теореме 6:

$$M\left[(X-MX)\cdot(Y-MY)\right]=M(X-MX)\cdot M(Y-MY),$$

а последнее выражение равно нулю, так как

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

Следовательно,

$$D(X+Y)=M(X-MX)^2+M(Y-MY)^2=DX+DY.$$

□

Обобщение свойства 4. Если случайные величины $X_1,X_2,...,X_n$ ***попарно независимы***, то справедлива формула:

$$D\left(X_1+X_2+...+X_n\right)=DX_1+DX_2+...+DX_n. \tag{25}$$

Доказательство провести самостоятельно, используя формулу:

$$\left(\sum_{k=1}^{n}a_k\right)^2=\sum_{k=1}^{n}a_k^2+2\sum_{1\le i<j\le n}^{n}a_i a_j.$$

Пример 23 (продолжение примеров 19 и 21). Для с.в. X, равной числу очков, выпавших на игральной кости, в примерах 19 и 21 подсчитано: $MX=\frac{7}{2}$, $M\left(X^2\right)=\frac{91}{6}$. Поэтому $DX=M\left(X^2\right)-(MX)^2=\frac{91}{6}-\frac{49}{4}=\frac{35}{12}$.

□

Пример 24 (продолжение примеров 18 и 20). Вычислим $D\mu_n$. Воспользуемся равенством (15):

$$\mu_n=I_{A_1^1}+I_{A_2^1}+...+I_{A_n^1}.$$

Точно так же, как это сделано в лемме 3, можно показать, что с.в. $I_{A_1^1},I_{A_2^1},...,I_{A_n^1}$ независимы в совокупности. Применяя формулу (25), получаем:

$$D\mu_n=DI_{A_1^1}+DI_{A_2^1}+...+DI_{A_n^1}.$$

Вычислим $DI_{A_k^1}$. Из свойства 2 математического ожидания следует, что $MI_{A_k^1}=P\left(A_k^1\right)=p$. Так как (по свойству 2 индикаторов) $I_{A_k^1}^2=I_{A_k^1}$, то и $M\left(I_{A_k^1}^2\right)=p$. Поэтому

$$DI_{A_k^1}=M\left(I_{A_k^1}^2\right)-\left[M\left(I_{A_k^1}\right)\right]^2=p-p^2=p(1-p)=pq.$$

То есть $D\mu_n=npq$.

□

ЧАСТЬ 2: СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Лекция 6

Случайные величины на счетном вероятностном пространстве

Определение 16. ***Счетным вероятностным пространством*** называется такое вероятностное пространство $(\Omega, \mathcal{F}, P)$ (см. определение 4), у которого элементы Ω можно занумеровать целыми неотрицательными числами (то есть $\Omega=\{\omega_0,\omega_1,\omega_2,...,\omega_n,...\}$ представляет собой некоторую последовательность), а σ-алгебра F представляет собой совокупность ***всех*** подмножеств множества Ω.

□

С.в. на счетном вероятностном пространстве определяется точно так же, как с.в. на конечном вероятностном пространстве (см. определение 10).

Определение 17. Пусть $X=X(\omega)$ – с.в. на счетном вероятностном пространстве $(\Omega, \mathcal{F}, P)$ и $x_0, x_1, x_2,..., x_m,...$ – всевозможные ***различные*** значения этой с.в. ***Законом распределения*** данной с.в. называется таблица с бесконечным числом столбцов:

X	x_0	x_1	x_2	...	x_m	...
P	p_0	p_1	p_2	...	p_m	...

где нижний ряд состоит из чисел p_k, равных вероятностям принятия случайной величиной $X=X(\omega)$ значений x_k, то есть $p_k=P(\omega\in\Omega\colon X(\omega)=x_k)$, $k=0,1,2,...,m,...$.

□

Заметим, что так как совокупность $x_0, x_1, x_2,..., x_m,...$ ***исчерпывает*** все значения, принимаемые с.в. X, то сумма чисел, стоящих в нижней строке таблицы, должна равняться единице, то есть $\sum_{k=0}^{\infty} p_k=1$ (напомним, что выражение $\sum_{k=0}^{\infty} p_k=p_0+p_1+p_2+...+p_k+...$ называется числовым рядом, сумма которого определяется формулой: $\sum_{k=0}^{\infty} p_k=\lim_{n\to\infty}\sum_{k=0}^{n} p_k$). Отметим также, что элементарные события, значения случайной величины и соответствующие вероятности индексируются нами, ***начиная с нуля***, только лишь для единообразия при рассмотрении дальнейших примеров.

Пример 25. Находясь в рамках схемы Бернулли, будем последовательно проводить испытания до тех пор, пока не достигнем первой неудачи. После этого испытания прекращаются. Очевидно, в качестве элементарных событий, соответствующих данному эксперименту, можно взять следующие вектора: $\omega_1=(0), \omega_2=(1,0), \omega_3=(1,1,0), \omega_4=(1,1,1,0),...$, причем к этой совокупности нужно добавить бесконечномерный вектор $\omega_0=(1,1,1,...,1,...)$, соответствующий такому (немыслимому) исходу, когда все время получаются успехи.

Применяя введенные ранее обозначения (см. раздел "Схема Бернулли"), получаем следующее представление для ω_k при $k\geq 1$: $\omega_k=A_1^1\cdot A_2^1\cdot...\cdot A_{k-1}^1\cdot A_k^0$. Так как события $A_1^1, A_2^1,...,A_{k-1}^1, A_k^0$ независимы в совокупности, то

$$P\{\omega_k\}=P\left(A_1^1\cdot A_2^1\cdot...\cdot A_{k-1}^1\cdot A_k^0\right)=P\left(A_1^1\right)\cdot P\left(A_2^1\right)\cdot...\cdot P\left(A_{k-1}^1\right)\cdot P\left(A_k^0\right)=p^{k-1}\cdot q.$$

Вычислим сумму вероятностей элементарных событий ω_k, начиная с $k=1$. Имеем:

$$\sum_{k=1}^{\infty} P\{\omega_k\}=q+pq+p^2q+p^3q+...=q\left(1+p+p^2+p^3+...\right).$$

В скобках стоит сумма известной из школьного курса бесконечно убывающей геометрической прогрессии. Эта сумма равна $\frac{1}{1-p}=\frac{1}{q}$. Таким образом, $\sum_{k=1}^{\infty} P\{\omega_k\}=q\cdot\frac{1}{q}=1$. С другой стороны, по аксиоме σ-аддитивности (см. определение 4) $\sum_{k=0}^{\infty} P\{\omega_k\}=1$. Следовательно, $P\{\omega_0\}=0$.

Рассмотрим с.в. X, равную числу успехов до первой неудачи. Имеем:

$$X(\omega_0)=1+1+1+...+1+...=+\infty,\ X(\omega_1)=0,\ X(\omega_2)=1,\ X(\omega_3)=2,...,X(\omega_{m+1})=m,....$$

Обозначим $x_0=+\infty, x_1=0, x_2=1, x_3=2,...,x_m=m-1, x_{m+1}=m,...$. Тогда:

$$p_0=P(\omega\in\Omega\colon X(\omega)=+\infty)=P\{\omega_0\}=0,\ p_1=P(\omega\in\Omega\colon X(\omega)=0)=P\{\omega_1\}=q,...,$$

$$p_m=P(\omega\in\Omega\colon X(\omega)=m-1)=P\{\omega_m\}=p^{m-1}q,\ p_{m+1}=P(\omega\in\Omega\colon X(\omega)=m)=P\{\omega_{m+1}\}=p^m q,...$$

Закон распределения с.в. X имеет вид:

X	$+\infty$	0	1	2	...	m	...
P	0	q	pq	p^2q	...	p^mq	...

В этом примере мы впервые столкнулись со случайной величиной, принимающей бесконечное значение. Вообще говоря, с такими с.в. работать не очень приятно, однако в данном случае нас спасает то обстоятельство, что вероятность принятия случайной величиной X значения $+\infty$ равна нулю. Так как событиями нулевой вероятности можно довольно часто пренебрегать, то мы можем в данной модели вместо пространства элементарных событий Ω рассмотреть "подчищенное" пространство $\widetilde{\Omega}=\{\omega_1,\omega_2,\omega_3,...\}$. Если ту же случайную величину рассмотреть на $\widetilde{\Omega}$, то ее закон распределения будет иметь вид:

X	0	1	2	...	m	...
P	q	pq	p^2q	...	p^mq	...

(26)

Полученная математическая модель столь же адекватна нашему эксперименту, как и первоначальная. Закон распределения (26) принято называть ***геометрическим распределением.***

□

<u>Пример 26</u>. На некотором счетном вероятностном пространстве $(\Omega,\mathcal{F},P)$ рассмотрим с.в. X, принимающую значения 0,1,2,...,m,... с вероятностями

$$p_m=P(\omega\in\Omega\colon X(\omega)=m)=\frac{a^m}{m!}\cdot e^{-a},$$

где число $a>0$. Из теории рядов хорошо известно, что $\sum_{m=0}^{\infty}\frac{a^m}{m!}=e^a$. Поэтому

$$\sum_{m=0}^{\infty}p_m=\sum_{m=0}^{\infty}\frac{a^m}{m!}\cdot e^{-a}=e^{-a}\cdot\sum_{m=0}^{\infty}\frac{a^m}{m!}=e^{-a}\cdot e^{a}=1.$$

Закон распределения с.в. X выражается таблицей:

X	0	1	2	...	m	...
P	e^{-a}	$a\cdot e^{-a}$	$\frac{a^2\cdot e^{-a}}{2}$	...	$\frac{a^m\cdot e^{-a}}{m!}$	...

(27)

Такой закон распределения называется ***пуассоновским*** (с параметром a).

□

Математическое ожидание случайной величины в случае счетного вероятностного пространства

В случае конечного вероятностного пространства мы видели, что любая с.в., принимающая конечные значения, обладает конечным математическим ожиданием. В случае счетного вероятностного пространства это далеко не так. Поэтому в данном разделе мы сузим число рассматриваемых с.в.: мы будем рассматривать только те с.в. X на счетном вероятностном пространстве $(\Omega, \mathcal{F}, P)$ (см. определение 16), для которых

$$\sum_{k=0}^{\infty} \left| X(\omega_k) \right| \cdot P\{\omega_k\} < \infty . \tag{28}$$

Если это условие не выполняется, то говорят, что с.в. X не имеет математического ожидания.

<u>*Определение 18*</u>. Математическим ожиданием с.в. X на счетном вероятностном пространстве $(\Omega, \mathcal{F}, P)$ называется число:

$$MX = \sum_{k=0}^{\infty} X(\omega_k) \cdot P\{\omega_k\} . \tag{29}$$

□

Из теории рядов известно, что если выполняется неравенство (28), то ряд в формуле (29) сходится, то есть MX есть некоторое действительное число.

Свойства 1, 2 и 3 математического ожидания для счетного вероятностного пространства формулируются и доказываются точно так же, как и соответствующие свойства матожидания с.в. на конечном вероятностном пространстве. Свойство 4 формулируется так: если с.в. X обладает законом распределения, выписанным в определении 17, то

$$MX = \sum_{k=0}^{\infty} x_k \cdot p_k . \tag{30}$$

Доказательство формулы (30) опускается.

Формулы (16) и (17), рассмотренные нами в случае конечного вероятностного пространства, принимают следующий вид:

$$Mf(X) = \sum_{k=0}^{\infty} f(x_k) \cdot p_k \tag{31}$$

и

$$Mg(X,Y) = \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} g(x_k, y_i) \cdot P(X = x_k,\, Y = y_i) . \tag{32}$$

Для справедливости этих формул достаточно предположить, что ряды, фигурирующие в формулах (31) и (32), абсолютно сходятся.

Определение независимых с.в. в случае счетного вероятностного пространства повторяет определение 14, только k и i нужно изменять от 0 до бесконечности, а теорема 6 полностью сохраняет силу.

Дисперсия случайной величины в случае счетного вероятностного пространства

Дисперсия с.в. X на счетном вероятностном пространстве $(\Omega, \mathcal{F}, P)$ определяется формулой (20) (см. определение 15). Она может быть записана в виде ряда с положительными членами:

$$DX = \sum_{k=0}^{\infty} \left(X(\omega_k) - MX \right)^2 \cdot P\{\omega_k\} . \tag{33}$$

Если сумма ряда (33) конечна, то говорят, что дисперсия существует и конечна. Если ряд (33) расходится (то есть его сумма равна $+\infty$), то говорят, что дисперсия не существует.

Свойства 1, 2 , 3 и 4 дисперсии для счетного вероятностного пространства формулируются и доказываются точно так же, как и соответствующие свойства дисперсии с.в. на конечном вероятностном пространстве.

Пример 27 (продолжение примера 25). Используя некоторые простые свойства рядов, можно доказать следующие формулы для математического ожидания и дисперсии с.в. X, введенной в примере 25:

$$MX=\sum_{k=0}^{\infty}kp^k q=\frac{p}{q}, \tag{34}$$

$$DX=\sum_{k=0}^{\infty}k^2 p^k q-\left(\frac{p}{q}\right)^2=\frac{p+p^2}{q^2}-\frac{p^2}{q^2}=\frac{p}{q^2}. \tag{35}$$

□

Пример 28 (продолжение примера 26). Вычислим более подробно математическое ожидание и дисперсию с.в. X, обладающей пуассоновским законом распределения (см. таблицу (27)). Применяя формулу (30), получаем:

$$MX=\sum_{k=0}^{\infty}x_k p_k=\sum_{k=0}^{\infty}k\cdot\frac{a^k\cdot e^{-a}}{k!}=\sum_{k=1}^{\infty}k\cdot\frac{a^k\cdot e^{-a}}{k!}=a\cdot\sum_{k=1}^{\infty}\frac{a^{k-1}\cdot e^{-a}}{(k-1)!}.$$

В последней сумме сделаем замену $m=k-1$ и продолжим вычисления:

$$MX=a\cdot\sum_{k=1}^{\infty}\frac{a^{k-1}\cdot e^{-a}}{(k-1)!}=a\cdot\sum_{m=0}^{\infty}\frac{a^m\cdot e^{-a}}{m!}=a,$$

так как $\sum_{m=0}^{\infty}\frac{a^m\cdot e^{-a}}{m!}=1$ (см. вычисления, проведенные в примере 26).

С целью вычислить DX, подсчитаем сначала $M\left(X^2\right)$. Представим с.в. X^2 в виде $X^2=X(X-1)+X$ и начнем вычисления, используя свойство линейности матожидания:

$$M\left(X^2\right)=M\left[X(X-1)+X\right]=M\left[X(X-1)\right]+MX=M\left[X(X-1)\right]+a.$$

Введя в рассмотрение функцию $f(x)=x(x-1)$, применим формулу (31):

$$M\left[X(X-1)\right]=M\left[f(X)\right]=\sum_{k=0}^{\infty}f\left(x_k\right)\cdot p_k=\sum_{k=0}^{\infty}k(k-1)\cdot\frac{a^k\cdot e^{-a}}{k!}=\sum_{k=2}^{\infty}k(k-1)\cdot\frac{a^k\cdot e^{-a}}{k!}=a^2\cdot\sum_{k=2}^{\infty}\frac{a^{k-2}\cdot e^{-a}}{(k-2)!}.$$

В последней сумме сделаем замену $m=k-2$ и продолжим вычисления:

$$M\left[X(X-1)\right]=a^2\cdot\sum_{k=2}^{\infty}\frac{a^{k-2}\cdot e^{-a}}{(k-2)!}=a^2\cdot\sum_{m=0}^{\infty}\frac{a^m\cdot e^{-a}}{m!}=a^2.$$

Таким образом, $M\left(X^2\right)=a^2+a$.

Применяя формулу (21) для вычисления дисперсии, получаем:

$$DX=M\left(X^2\right)-(MX)^2=a^2+a-a^2=a.$$

Итак, мы доказали, что для с.в. X, распределенной по пуассоновскому закону с параметром a, справедливы соотношения:

$$MX=DX=a. \tag{36}$$

□

ЧАСТЬ 2: СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Лекция 7

Случайные величины на произвольном вероятностном пространстве

Начиная с этого момента, мы будем работать на некотором вероятностном пространстве $(\Omega, \mathcal{F}, P)$ (см. определение 4), а через $\hat{\mathcal{F}}$ будем обозначать борелевскую σ-алгебру на действительной прямой *R* (см. пример 5 и его продолжения). Напомним, что $\hat{\mathcal{F}}$ есть наименьшая σ-алгебра, содержащая все интервалы вида $[a, b)$ (включая интервалы $(-\infty, b)$ и $[a, +\infty)$).

Определение 19. Случайной величиной X на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ называется такая функция $X = X(\omega)$ от элементарного события $\omega \in \Omega$, что для любых чисел $a < b$ событие $\{\omega \in \Omega: a \le X(\omega) < b\}$ содержится в σ-алгебре $\mathcal{F}$.

□

Можно доказать (это выходит за рамки нашей программы), что если *X* – с.в. на $(\Omega, \mathcal{F}, P)$, то для любого подмножества $U \in \hat{\mathcal{F}}$ событие $\{\omega \in \Omega: X(\omega) \in U\}$ также содержится в σ-алгебре $\mathcal{F}$. В частности, событие $\{\omega \in \Omega: X(\omega) \in (-\infty, x)\} = \{\omega \in \Omega: X(\omega) < x\}$ содержится в $\mathcal{F}$ при любом действительном *x*.

Определение 20. ***Функцией распределения*** случайной величины $X = X(\omega)$ называется функция $y = F(x)$ с областью определения *R*, задающаяся формулой:

$$F(x) = P\{\omega \in \Omega: X(\omega) < x\}. \tag{37}$$

□

Формулу (37) часто записывают в более лаконичном виде:

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

Иногда, чтобы подчеркнуть, что рассматриваемая функция распределения есть функция распределения именно случайной величины *X*, мы употребляем обозначение $F_X(x)$).

Свойства функции распределения

1. $0 \le F(x) \le 1$ для всех $x \in R$.
2. Функция $y = F(x)$ монотонно возрастает, то есть $x_1 \le x_2 \Rightarrow F(x_1) \le F(x_2)$.
3. $F(-\infty) := \lim\limits_{x \to -\infty} F(x) = 0, \; F(+\infty) := \lim\limits_{x \to +\infty} F(x) = 1$.
4. Функция $y = F(x)$ непрерывна слева, то есть для любого $a \in R$ $\lim\limits_{x \to a-0} F(x) = F(a)$.

Для доказательства некоторых из этих свойств нам понадобится следующая лемма, представляющая и самостоятельный интерес.

Лемма 5. Пусть $\{A_k\}_{k=1}^{\infty}$ – семейство событий на вероятностном пространстве $(\Omega, \mathcal{F}, P)$. Тогда:

1) если $A_1 \supset A_2 \supset \ldots \supset A_k \supset \ldots$ и $A = \bigcap\limits_{k=1}^{\infty} A_k$, то $P(A) = \lim\limits_{k \to \infty} P(A_k)$;

2) если $A_1 \subset A_2 \subset \ldots \subset A_k \subset \ldots$ и $A = \bigcup\limits_{k=1}^{\infty} A_k$, то $P(A) = \lim\limits_{k \to \infty} P(A_k)$.

Доказательство опускается.

□

<u>*Доказательство*</u> свойств функции распределения.

1. Так как вероятность любого события не меньше нуля и не больше единицы, то данное свойство непосредственно следует из формулы (37).

2. Если $x_1 \le x_2$, то $\{\omega \in \Omega\colon X(\omega) < x_1\} \subset \{\omega \in \Omega\colon X(\omega) < x_2\}$. Применяя свойство 3 вероятности, получаем: $P\{\omega \in \Omega\colon X(\omega) < x_1\} \le P\{\omega \in \Omega\colon X(\omega) < x_2\}$. Но левая часть последнего неравенства – это $F(x_1)$, а правая – $F(x_2)$. Следовательно, $F(x_1) \le F(x_2)$.

3. Пусть последовательность $\{x_k\}_{k=1}^{\infty}$ строго монотонно убывает к $-\infty$, то есть $x_1 > x_2 > \ldots > x_k > \ldots$ и $\lim\limits_{k\to\infty} x_k = -\infty$. Обозначим $A_k = \{\omega \in \Omega\colon X(\omega) < x_k\}$. Легко видеть, что $A_1 \supset A_2 \supset \ldots \supset A_k \supset \ldots$ и $A = \bigcap\limits_{k=1}^{\infty} A_k = \varnothing$. (Последнее соотношение доказывается от противного. Пусть существует $\omega \in A$. Так как $\lim\limits_{k\to\infty} x_k = -\infty$, то найдется номер n такой, что $x_n < X(\omega)$. Следовательно, $\omega \notin A_n$. Но это противоречит тому, что $A = \bigcap\limits_{k=1}^{\infty} A_k$, то есть ω должен принадлежать каждому сомножителю, в том числе и A_n.) Применяя пункт 1) леммы 5, получаем, что $0 = P(A) = \lim\limits_{k\to\infty} P(A_k) = \lim\limits_{k\to\infty} F(x_k)$. По определению предела функции, в силу произвольности $\{x_k\}_{k=1}^{\infty}$ это как раз и означает, что $\lim\limits_{x\to-\infty} F(x) = 0$.

Соотношение $\lim\limits_{x\to+\infty} F(x) = 1$ доказывается аналогично (проделайте это!).

4. Пусть последовательность $\{x_k\}_{k=1}^{\infty}$ строго монотонно возрастает к a, то есть $x_1 < x_2 < \ldots < x_k < \ldots$ и $\lim\limits_{k\to\infty} x_k = a$. Обозначим $A_k = \{\omega \in \Omega\colon X(\omega) < x_k\}$. Легко видеть, что $A_1 \subset A_2 \subset \ldots \subset A_k \subset \ldots$ и $A = \bigcup\limits_{k=1}^{\infty} A_k = \{\omega \in \Omega\colon X(\omega) < a\}$ (см. рассуждения, проведенные в доказательстве предыдущего свойства). Применяя пункт 2) леммы 5, получаем, что $F(a) = P\{\omega \in \Omega\colon X(\omega) < a\} = P(A) = \lim\limits_{k\to\infty} P(A_k) = \lim\limits_{k\to\infty} F(x_k)$. По определению предела функции, в силу произвольности $\{x_k\}_{k=1}^{\infty}$ это означает, что $\lim\limits_{x\to a-0} F(x) = F(a)$.

□

<u>*Теорема 7*</u>. Пусть $X = X(\omega)$ – с.в. на $(\Omega, \mathcal{F}, P)$. Тогда справедливы формулы:

$$P(a \le X < b) = F(b) - F(a) \tag{38}$$

и

$$P(X = a) = F(a+0) - F(a), \tag{39}$$

где $F(a+0) := \lim\limits_{x\to a+0} F(x)$, а a и b – произвольные действительные числа.

<u>*Доказательство*</u>. Так как $\{X < b\} = \{X < a\} + \{a \le X < b\}$ и слагаемые в этой сумме несовместны, то в силу аксиомы аддитивности $P(X < b) = P(X < a) + P(a \le X < b)$,то есть $F(b) = F(a) + P(a \le X < b)$, откуда немедленно следует (38).

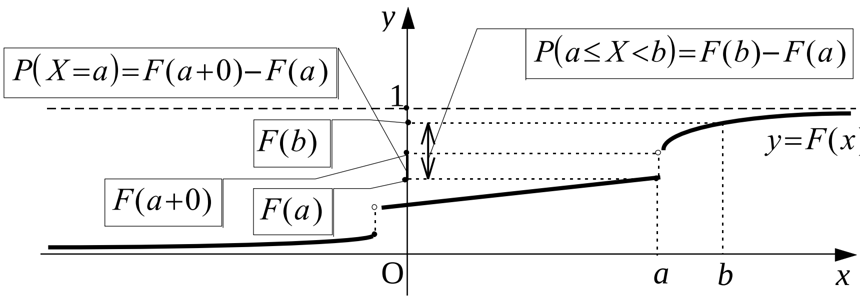
Пусть последовательность $\{x_k\}_{k=1}^{\infty}$ строго монотонно убывает к a, то есть $x_1 > x_2 > \ldots > x_k > \ldots$ и $\lim\limits_{k\to\infty} x_k = a$. Обозначим $A_k = \{\omega \in \Omega\colon X(\omega) < x_k\}$. Легко видеть, что $A_1 \supset A_2 \supset \ldots \supset A_k \supset \ldots$ и $A = \bigcap\limits_{k=1}^{\infty} A_k = \{\omega \in \Omega\colon X(\omega) \le a\}$. Применяя пункт 1) леммы 5, имеем:

$P(X \le a) = P(A) = \lim\limits_{k\to\infty} P(A_k) = \lim\limits_{k\to\infty} F(x_k)$. По определению предела функции это означает, что $P(X \le a) = \lim\limits_{x\to a+0} F(x) = F(a+0)$. Но $P(X \le a) = P(X < a) + P(X = a) = F(a) + P(X = a)$ вследствие аксиомы аддитивности. Следовательно, $F(a) + P(X = a) = F(a+0)$, откуда вытекает (39).

□

В случае конечного и счетного вероятностных пространств мы характеризовали случайную величину с помощью таблицы, которую называли законом распределения с.в. (см. определения 12 и 17). Нижние строки этих таблиц состояли из вероятностей $P(X = x_k)$, где через x_k мы обозначали возможные значения с.в. *X*. Формула (39) показывает, что в общем случае такой способ описания случайных величин неприемлем. Действительно, если функция $y = F(x)$ непрерывна (примеры таких функций распределения будут вскоре нами построены), то для любого действительного a получаем $F(a+0) = F(a)$, и из формулы (39) следует $P(X = a) = 0$, что противоречит принципу построения таблиц распределений в случае конечного и счетного вероятностных пространств.

В случае общего вероятностного пространства функция распределения выполняет роль рассмотренных выше таблиц распределения. Ввиду свойств 1–4 функции распределения, примерный вид ее графика таков:



Важную роль играет следующая теорема, которая позволяет строить случайные величины по заданным функциям распределения.

<u>*Теорема 8*</u>. Пусть функция $y = F(x)$, определенная на всей действительной прямой, удовлетворяет условиям 1-4 функции распределения. Тогда существует вероятностное пространство и на нем случайная величина *X*, чья функция распределения совпадает с $F(x)$.

<u>*Доказательство*</u>. В качестве пространства элементарных событий возьмем действительную прямую *R*, а в качестве σ-алгебры событий выберем борелевскую σ-алгебру $\hat{\mathfrak{F}}$. Определим функцию множеств $\hat{P}$ на интервалах $[a, b)$ следующим образом:

$$\hat{P}\{[a,b)\} = F(b) - F(a). \tag{40}$$

Можно доказать (это выходит за рамки нашей программы), что функция множеств $\hat{P}$ единственным образом доопределяется на всех остальных подмножествах из σ-алгебры $\hat{\mathfrak{F}}$ так, что $\hat{P}$ становится вероятностью на $\hat{\mathfrak{F}}$ (то есть будут выполнены все аксиомы из определения 4). Таким образом, мы построили вероятностное пространство $(R, \hat{\mathfrak{F}}, \hat{P})$.

Определим на построенном $(R, \hat{\mathfrak{F}}, \hat{P})$ тождественную функцию $X(\omega) = \omega$, $\omega \in R$. Легко видеть, что *X* – с.в. на $(R, \hat{F}, \hat{P})$ (докажите это!). Вычислим ее функцию распределения $F_X(x)$. Имеем:

$$F_X(x) = \hat{P}\{\omega \in R \colon X(\omega) < x\} = \hat{P}\{\omega \in R \colon \omega < x\} = \hat{P}\{(-\infty, x)\}.$$

Обозначим $A_k=\left[-k,x\right)$. Тогда $A_1\subset A_2\subset\ldots\subset A_k\subset\ldots$ и $A=\bigcup_{k=1}^{\infty}A_k=\left(-\infty,x\right)$. По пункту 2) леммы 5 имеем, используя формулу (40):

$$\hat{P}\left(-\infty,x\right)=\hat{P}(A)=\lim_{k\to\infty}\hat{P}\left(A_k\right)=\lim_{k\to\infty}\hat{P}\left(A_k\right)=\lim_{k\to\infty}\left(F(x)-F(-k)\right)=F(x)-\lim_{k\to\infty}F(-k).$$

Так как функция $y=F(x)$ удовлетворяет свойству 3 функции распределения, то $\lim_{k\to\infty}F(-k)=F(-\infty)=0$. Таким образом, $\hat{P}\left(-\infty,x\right)=F(x)$ и, следовательно, $F_X(x)=F(x)$.

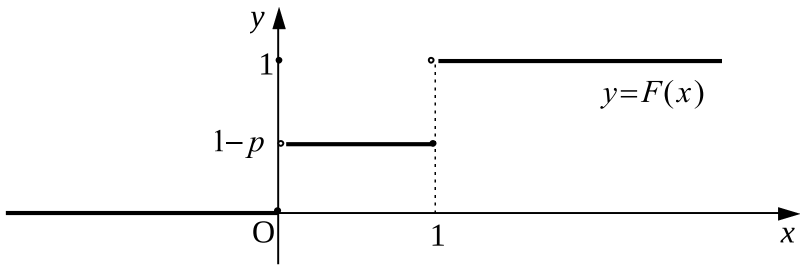
□

Заметим, что примененные в доказательстве теоремы 8 вероятностное пространство $\left(R,\hat{\mathcal{F}},\hat{P}\right)$ и тождественная с.в. X на нем называются ***канонической реализацией*** случайной величины по заданной функции распределения.

<u>*Пример 29*</u>. Вычислим функцию распределения $F(x)$ индикатора I_A (определение 11 остается действующим и в случае произвольного вероятностного пространства), если $P(A)=p$. Имеем:

$$F(x)=P\left(I_A<x\right)=\begin{cases} P(\varnothing)=0, \textit{ если } x\le 0,\\ P(\overline{A})=1-p, \textit{ если } 0<x\le 1,\\ P(\Omega)=1, \textit{ если } x>1.\end{cases}$$

График функции $y=F(x)$ имеет вид:



Он представляет собой так называемую кусочно-постоянную функцию с двумя скачками. Первый скачок график испытывает в точке $x_1=0$, причем величина скачка равна $1-p$, то есть вероятности принятия индикатором I_A значения 0. Второй скачок график испытывает в точке $x_2=1$, причем величина скачка равна p, то есть вероятности принятия индикатором I_A значения 1. Теперь совершенно ясно, как с помощью этой функции распределения составить закон распределения (таблицу) из примера 17 и наоборот.

□

<u>*Определение 21*</u>. Если с вероятностью 1 с.в. X принимает конечное или счетное (записывающееся в последовательность) число значений, то такая случайная величина называется дискретной.

□

Все случайные величины, определенные на конечных и счетных вероятностных пространствах являются дискретными. Обратно, каждую дискретную случайную величину можно реализовать на конечном или счетном вероятностном пространстве (подумайте, как!).

ЧАСТЬ 2: СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Лекция 8

Непрерывные случайные величины

Определение 22. Случайную величину *X*, определенную на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, будем называть абсолютно непрерывной, если ее функцию распределения $F(x)$ можно представить в виде:

$$F(x)=\int_{-\infty}^{x} p(t)dt\,, \tag{41}$$

где $p(x)\geq 0$ – некоторая функция на числовой прямой, называемая плотностью распределения с.в. *X*.

□

Если нам нужно будет подчеркнуть, что функция $p(x)$ является плотностью распределения ***именно*** с.в. *X*, то мы ее будем обозначать через $p_X(x)$.

Из формулы (41) следует формула, выражающая плотность распределения $p(x)$ через функцию распределения $F(x)$:

$$p(x)=F^{/}(x)\,. \tag{42}$$

Очевидно, что формула (42) верна для тех точек $x\in R$, в которых функция $p(x)$ непрерывна (это следует из теоремы о производной от интеграла с переменным верхним пределом). На самом деле, это равенство справедливо для значительно большего числа точек $x\in R$, однако, вообще говоря, не для всех точек $x\in R$ (см. ниже пример 30).

Теорема 9. Неотрицательная на действительной прямой *R* функция $p(x)$ является плотностью распределения некоторой с.в. *X* тогда и только тогда, когда

$$\int_{-\infty}^{+\infty} p(x)dx=1\,. \tag{43}$$

Доказательство. Необходимость. Пусть $p(x)$ – плотность распределения с.в. *X*, то есть выполняется соотношение (41), где $F(x)$ – функция распределения с.в. *X*. Перейдя в равенстве (41) к пределу при $x\to+\infty$, получим:

$$\lim_{x\to+\infty} F(x)=\lim_{x\to+\infty}\int_{-\infty}^{x} p(t)dt\,.$$

По свойству 3 функции распределения имеем: $\lim\limits_{x\to+\infty} F(x)=F(+\infty)=1$. Следовательно, предел правой части существует и, по определению, равен несобственному интегралу $\int_{-\infty}^{+\infty} p(t)dt$.

Таким образом, получаем равенство $1=\int_{-\infty}^{+\infty} p(t)dt$, которое после замены *t* на *x* дает (43).

Достаточность. Пусть функция $p(x)\geq 0$ такова, что выполняется (43). Заменяя *x* новой переменной интегрирования *t*, получаем равенство $\int_{-\infty}^{+\infty} p(t)dt=1$. Обозначим $F(x)=\int_{-\infty}^{x} p(t)dt$ и докажем, что функция $F(x)$ удовлетворяет свойствам 1–4 функции распределения.

Так как $p(t)\geq 0$, то по свойству позитивности интеграла получаем $F(x)=\int\limits_{-\infty}^{x} p(t)dt\geq 0$.

Так как по свойству аддитивности интеграла $\int\limits_{-\infty}^{+\infty} p(t)dt=\int\limits_{-\infty}^{x} p(t)dt+\int\limits_{x}^{+\infty} p(t)dt$, то:

$$F(x)=\int\limits_{-\infty}^{x} p(t)dt=\int\limits_{-\infty}^{+\infty} p(t)dt-\int\limits_{x}^{+\infty} p(t)dt=1-\int\limits_{x}^{+\infty} p(t)dt\leq 1 \tag{44}$$

(последнее неравенство опять записано с учетом свойства позитивности). Таким образом, $0\leq F(x)\leq 1$, и свойство 1 функции распределения доказано.

Свойство 2 также доказывается с применением свойств позитивности и аддитивности интеграла: если $x_1\leq x_2$,то

$$F(x_1)=\int\limits_{-\infty}^{x_1} p(t)dt\leq \int\limits_{-\infty}^{x_1} p(t)dt+\int\limits_{x_1}^{x_2} p(t)dt=\int\limits_{-\infty}^{x_2} p(t)dt=F(x_2).$$

Из выкладок в цепочке (44) следует, что $F(x)=1-\int\limits_{x}^{+\infty} p(t)dt$. Переходя в этом равенстве к пределу при $x\to-\infty$, получаем:

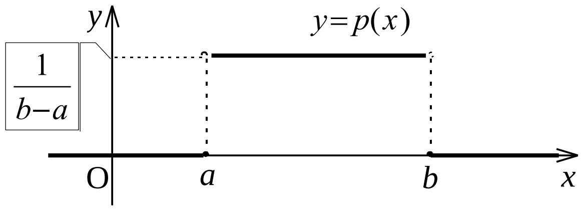
$$F(-\infty)=\lim_{x\to-\infty} F(x)=1-\lim_{x\to-\infty}\int\limits_{x}^{+\infty} p(t)dt=1-\int\limits_{-\infty}^{+\infty} p(t)dt=1-1=0.$$

Свойство $F(+\infty)=1$ получается переходом к пределу при $x\to+\infty$ в равенстве (41).

Из свойств интеграла с переменным верхним пределом следует, что функция $F(x)$ непрерывна на *R*, если $p(x)$ непрерывна на *R*. Можно доказать (это выходит за рамки нашей программы), что $F(x)$ непрерывна на *R* и без предположения о непрерывности $p(x)$. Тем более, $F(x)$ непрерывна слева на *R*. То есть свойство 4 функции распределения также выполнено.

Применяя теорему 8, заключаем, что существует вероятностное пространство и на нем случайная величина *X*, чья функция распределения совпадает с $F(x)$ и, следовательно, чья плотность совпадает с $p(x)$. □

<u>*Пример 30*</u> (***равномерное распределение*** на отрезке). Рассмотрим функцию:



$$y=p(x)=\begin{cases}\dfrac{1}{b-a}, & \text{если } a<x<b,\\ 0, & \text{если } x\leq a \text{ или } x\geq b,\end{cases}$$

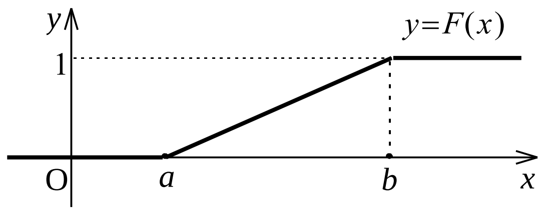
график которой представлен на рисунке. Так как $p(x)\geq 0$ и

$$\int\limits_{-\infty}^{+\infty} p(x)dx=\int\limits_{-\infty}^{a} p(x)dx+\int\limits_{a}^{b} p(x)dx+\int\limits_{b}^{+\infty} p(x)dx=\int\limits_{-\infty}^{a} 0dx+\int\limits_{a}^{b}\frac{1}{b-a}dx+\int\limits_{b}^{+\infty} 0dx=\frac{1}{b-a}\cdot(b-a)=1,$$

то по теореме 9 функция $p(x)$ является плотностью распределения некоторой с.в. *X*. Распределение этой с.в. называется ***равномерным распределением*** на отрезке $[a,b]$.

Используя формулу (41), вычислим функцию распределения $F(x)$ с.в. *X*. Если $x\leq a$, то $F(x)=\int\limits_{-\infty}^{x} p(t)dt=\int\limits_{-\infty}^{x} 0dt=0$. Если $a<x\leq b$, то $F(x)=\int\limits_{-\infty}^{a} p(t)dt+\int\limits_{a}^{x} p(t)dt=0+\int\limits_{a}^{x}\frac{1}{b-a}dt=\frac{x-a}{b-a}$.

Наконец, если $x>b$, то $F(x)=\int\limits_{-\infty}^{x} p(t)dt=\int\limits_{-\infty}^{a} p(t)dt+\int\limits_{a}^{b} p(t)dt+\int\limits_{b}^{x} p(t)dt=0+\int\limits_{a}^{b}\frac{1}{b-a}dt+0=1$. То есть:

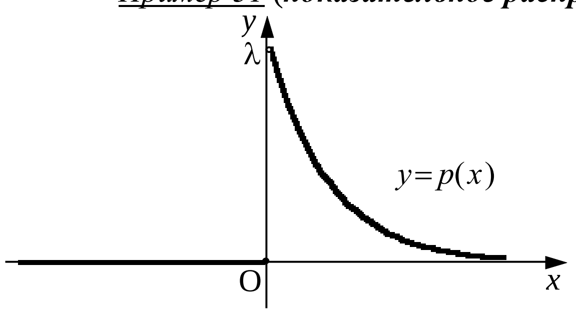


$$F(x)=\begin{cases} 0, & \text{если } x\le a,\\ \dfrac{x-a}{b-a}, & \text{если } a<x\le b,\\ 1, & \text{если } x>b,\end{cases}$$

(см. график этой функции на рисунке). Вид графика функции $y=F(x)$ дает нам возможность сделать заключение, что в этом случае равенство (42) не выполняется в точках $x=a$ и $x=b$ (в этих точках даже не существует производная $F^{/}(x)$). В остальных же точках равенство (42) выполнено как в точках непрерывности плотности $y=p(x)$.

□

<u>*Пример 31*</u> (***показательное распределение***). Рассмотрим функцию



$$y=p(x)=\begin{cases} \lambda e^{-\lambda x}, & \text{если } x>0,\\ 0, & \text{если } x\le 0,\end{cases}$$

график которой представлен на рисунке. Так как $p(x)\ge 0$ и

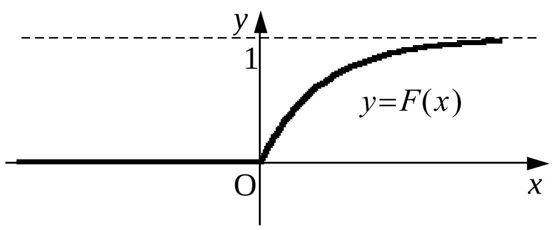
$$\int_{-\infty}^{+\infty} p(x)dx = \int_{-\infty}^{0} p(x)dx + \int_{0}^{+\infty} p(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{+\infty} \lambda e^{-\lambda x} dx =$$

$$=-e^{-\lambda x}\Big|_0^{+\infty}=-(0-1)=1,$$

то по теореме 9 функция $p(x)$ является плотностью распределения некоторой с.в. X. Распределение этой с.в. называется ***показательным распределением*** с параметром λ.

Используя формулу (41), вычислим функцию распределения $F(x)$ с.в. X. Если $x\le 0$, то $F(x)=\int_{-\infty}^{x} p(t)dt=\int_{-\infty}^{x} 0dt=0$. Если $0<x<+\infty$, то

$$F(x)=\int_{-\infty}^{0} p(t)dt+\int_{0}^{x} p(t)dt=0+\int_{0}^{x}\lambda e^{-\lambda t}dt=-e^{-\lambda t}\Big|_0^x=1-e^{-\lambda x}.$$



Таким образом:

$$F(x)=\begin{cases} 0, & \text{если } x\le 0,\\ 1-e^{-\lambda x}, & \text{если } x>0\end{cases}$$

(см. график этой функции на рисунке).

□

Перед тем, как рассмотреть следующий (особенно важный) пример, введем обозначения, которые будем систематически использовать в дальнейшем. Итак, обозначим:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ \Phi(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt,\ \Phi_0(x)=\int_{0}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt \qquad (45)$$

Функцию $\Phi(x)$ часто называют ***интегралом вероятности Гаусса***. Из теории несобственных интегралов известно следующее равенство:

$$\Phi(+\infty)=\int_{-\infty}^{+\infty}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=1. \qquad (46)$$

Ясно, что функция $\varphi(x)$ четная, то есть $\varphi(-x)=\varphi(x)$. Отсюда следует, что функция $\Phi_0(x)$ нечетная. Действительно:

$$\Phi_0(-x)=\int_0^{-x}\varphi(t)dt=\left|\begin{array}{lcr} z=-t & \Rightarrow & dt=-dz \\ t=0 & \Rightarrow & z=0 \\ t=-x & \Rightarrow & z=x \end{array}\right|=-\int_0^x\varphi(-z)dz=-\int_0^x\varphi(z)dz=-\Phi_0(x).$$

Таким образом:

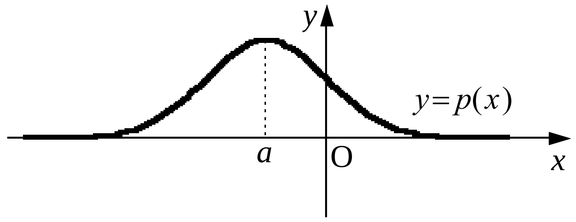
$$1=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-\frac{t^2}{2}}dt+\frac{1}{\sqrt{2\pi}}\int_{0}^{+\infty}e^{-\frac{t^2}{2}}dt=-\frac{1}{\sqrt{2\pi}}\int_{0}^{-\infty}e^{-\frac{t^2}{2}}dt+\Phi_0(+\infty)=-\Phi_0(-\infty)+\Phi_0(+\infty)=2\Phi_0(+\infty).$$

То есть, $\Phi_0(+\infty)=0{,}5$. Из проведенной выкладки также вытекает, что $\Phi(0)=0{,}5$ и

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-\frac{t^2}{2}}dt+\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt=\Phi(0)+\Phi_0(x)=\frac{1}{2}+\Phi_0(x). \qquad (47)$$

Таблицу значений функции $\Phi_0(x)$ можно найти в любом пособии по теории вероятностей.

Пример 32 (***нормальное распределение***). Рассмотрим функцию



$$p(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}=\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right),$$

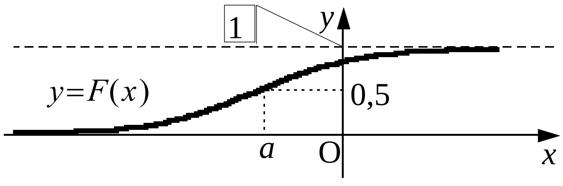
($a\in R, \sigma>0$), график которой представлен на рисунке. Этот график симметричен относительно прямой $x=a$ и имеет в точке $x=a$ единственный экстремум (максимум), равный $\frac{1}{\sqrt{2\pi}\cdot\sigma}$. С уменьшением σ этот график становится все более островершинным. Изменение *a* при постоянном σ не меняет форму графика, а вызывает лишь его смещение по оси абсцисс. Так как $p(x)\geq 0$ и

$$\int_{-\infty}^{+\infty}p(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\left|\begin{array}{lcr} t=\frac{x-a}{\sigma} & \Rightarrow & x=\sigma t+a \Rightarrow dx=\sigma dt \\ x=-\infty & \Rightarrow & t=-\infty \\ x=+\infty & \Rightarrow & t=+\infty \end{array}\right|=\int_{-\infty}^{+\infty}\varphi(t)dt=1,$$

то по теореме 9 функция $p(x)$ является плотностью распределения некоторой с.в. *X*. Распределение этой с.в. называется ***нормальным распределением*** с параметрами *a* и σ.

Используя формулу (41), вычислим функцию распределения $F(x)$ с.в. *X*. Имеем:

$$F(x)=\int_{-\infty}^{x}p(t)dt=\int_{-\infty}^{x}\frac{1}{\sigma}\varphi\left(\frac{t-a}{\sigma}\right)dt=\left|\begin{array}{lcr} z=\frac{t-a}{\sigma} & \Rightarrow & t=\sigma z+a \Rightarrow dt=\sigma dz \\ t=-\infty & \Rightarrow & z=-\infty \\ t=x & \Rightarrow & z=\frac{x-a}{\sigma} \end{array}\right|=\int_{-\infty}^{\frac{x-a}{\sigma}}\varphi(z)dz=\Phi\left(\frac{x-a}{\sigma}\right). \qquad (48)$$



На рисунке представлен график функции распределения нормальной с.в. с параметрами *a* и σ.

□

ЧАСТЬ 2: СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Лекция 9

Математическое ожидание абсолютно непрерывных случайных величин

<u>*Определение 23*</u>. Математическим ожиданием абсолютно непрерывной с.в. X с плотностью распределения $p(x)$ (см. определение 22) называется число

$$MX=\int_{-\infty}^{+\infty} xp(x)dx\,, \tag{49}$$

при условии, что несобственный интеграл абсолютно сходится (то есть, $\int_{-\infty}^{+\infty}|x|\cdot p(x)dx<+\infty$). В противном случае (то есть, если $\int_{-\infty}^{+\infty}|x|\cdot p(x)dx=+\infty$) говорят, что математическое ожидание не существует.

□

Заметим, что выражение (49) аналогично свойству 4 математического ожидания, доказанного нами для дискретных случайных величин (см. формулы (14) и (30)). В этих лекциях мы не можем дать определение матожидания произвольной случайной величины, так как это потребовало бы введения общей концепции интеграла (интеграл Лебега), что выходит за рамки программ технических вузов. Формула (49) в вычислительном плане хороша, однако в теоретическом плане она обладает существенными недостатками. Например, используя (49) невозможно доказать ***свойства линейности и монотонности математического ожидания*** (см раздел "Математическое ожидание в случае конечного вероятностного пространства"), которые, тем не менее, ***справедливы.*** Эти доказательства невозможны потому, что сумма абсолютно непрерывных с.в. может не иметь плотности распределения (приведите пример!).

<u>*Пример 33*</u> (продолжение примера 30). Вычислим математическое ожидание с.в. X, равномерно распределенной на отрезке $\left[a,b\right]$. Применяя формулу (49), имеем:

$$MX=\int_{-\infty}^{+\infty} xp(x)dx=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\frac{x^2}{2}\bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2}\,. \tag{50}$$

Таким образом, $MX=\dfrac{a+b}{2}$ находится на середине отрезка $\left[a,b\right]$.

□

<u>*Пример34*</u> (продолжение примера 31).). Вычислим математическое ожидание с.в. X, имеющей показательное распределение с параметром λ. Применяя формулу (49), имеем:

$$MX=\int_{-\infty}^{+\infty} xp(x)dx=\int_0^{+\infty} x\cdot\lambda e^{-\lambda x}dx=\left|\begin{matrix} u=x & \Rightarrow & du=dx \\ dv=\lambda e^{-\lambda x} & \Rightarrow & v=\int\lambda e^{-\lambda x}dx=-e^{-\lambda x}\end{matrix}\right|=-xe^{-\lambda x}\Big|_0^{+\infty}+\int_0^{+\infty} e^{-\lambda x}dx=$$

$$=-\frac{1}{\lambda}\cdot e^{-\lambda x}\bigg|_0^{+\infty}=0-\left(-\frac{1}{\lambda}\right)=\frac{1}{\lambda}\,.$$

Формула

$$MX=\frac{1}{\lambda} \tag{51}$$

интуитивно ясна. Действительно, чем больше λ, тем плотнее правая часть графика функции $y=p(x)$ примыкает к оси ординат (см. рисунок в примере 31). А это означает, что почти вся "масса" показательного распределения (с этим параметром λ) сосредотачивается вблизи нуля. То есть математическое ожидание должно быть малым.

□

Пример 35 (продолжение примера 32). Вычислим математическое ожидание с.в. X, имеющей нормальное распределение с параметрами a и σ. Применяя формулу (49), имеем:

$$MX=\int_{-\infty}^{+\infty} xp(x)dx=\int_{-\infty}^{+\infty} x\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}(x-a+a)\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}\frac{x-a}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx+a\cdot\int_{-\infty}^{+\infty}\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=$$

$$=\int_{-\infty}^{+\infty}\frac{x-a}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx+a.$$

В последнем интеграле сделаем замену переменных:

$$\int_{-\infty}^{+\infty}\frac{x-a}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\left|\begin{array}{lcr} t=\dfrac{x-a}{\sigma} & \Rightarrow & x=\sigma t+a\Rightarrow dx=\sigma dt\\ x=-\infty & \Rightarrow & t=-\infty\\ x=+\infty & \Rightarrow & t=+\infty\end{array}\right|=\sigma\cdot\int_{-\infty}^{+\infty}t\varphi(t)dt.$$

Но так как подынтегральная функция $t\varphi(t)$ – нечетная, а интеграл берется по множеству, симметричному относительно нуля, то этот интеграл равен нулю (читателю предлагается получить нуль прямым вычислением, а именно, применить формулу интегрирования по частям). Таким образом,

$$MX=a. \qquad (52)$$

Этот результат естественен в силу того, что прямая $x=a$ является осью симметрии графика функции $y=p(x)$ (см. рисунок в примере 32).

□

Теоретическое упражнение. Доказать следующее общее утверждение: если какая-то с.в. X имеет плотность распределения $y=p(x)$, график которой симметричен относительно прямой $x=a$, и математическое ожидание существует, то $MX=a$.

Для абсолютно непрерывных с.в. аналог теоремы 4 (см. также формулу (31)) имеет вид:

$$Mf(X)=\int_{-\infty}^{+\infty}f(x)\cdot p(x)dx, \qquad (53)$$

где $p(x)$ есть плотность распределения с.в. X, а несобственный интеграл справа абсолютно сходится. Доказательство формулы (53) опускается.

Пример 36 (продолжение примеров 30 и 33). Для с.в. X, равномерно распределенной на отрезке $[a,b]$, найдем $M(X^2)$. Введя функцию $f(x)=x^2$ и применяя формулу (53), имеем:

$$M(X^2)=\int_{-\infty}^{+\infty}x^2\cdot p(x)dx=\int_a^b x^2\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\frac{x^3}{3}\bigg|_a^b=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3}. \qquad (54)$$

□

Пример 37 (продолжение примеров 31 и 34). Для с.в. X, имеющей показательное распределение с параметром λ, найдем $M(X^2)$. Введя функцию $f(x)=x^2$ и применяя формулу (53), имеем:

$$M(X^2)=\int_{-\infty}^{+\infty}x^2\cdot p(x)dx=\int_0^{+\infty}x^2\cdot\lambda e^{-\lambda x}dx=$$

$$=\left|\begin{array}{lcr} u=x^2 & \Rightarrow & du=2xdx\\ dv=\lambda e^{-\lambda x}dx & \Rightarrow & v=\int\lambda e^{-\lambda x}dx=-e^{-\lambda x}\end{array}\right|=-x^2e^{-\lambda x}\Big|_0^{+\infty}+2\int_0^{+\infty}xe^{-\lambda x}dx=\frac{2}{\lambda}\cdot\int_0^{+\infty}x\cdot\lambda e^{-\lambda x}dx=\frac{2}{\lambda}\cdot MX=\frac{2}{\lambda^2}.$$

Итак,

$$M\left(X^2\right)=\frac{2}{\lambda^2}. \qquad (55)$$

□

<u>*Пример 38*</u> (продолжение примеров 32 и 35). Для с.в. *X*, имеющей нормальное распределение с параметрами *a* и σ, найдем $M\left(X^2\right)$. Введя функцию $f(x)=x^2$ и применяя формулу (53), имеем:

$$M\left(X^2\right)=\int_{-\infty}^{+\infty}x^2\cdot p(x)dx=\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}\left[(x-a)+a\right]^2\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}(x-a)^2\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx+$$

$$+\int_{-\infty}^{+\infty}2a(x-a)\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx+\int_{-\infty}^{+\infty}a^2\cdot\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right)dx=\int_{-\infty}^{+\infty}\sigma\cdot\left(\frac{x-a}{\sigma}\right)^2\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx+0+a^2=$$

$$=\int_{-\infty}^{+\infty}\sigma\cdot\left(\frac{x-a}{\sigma}\right)^2\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx+a^2$$

(мы использовали результаты вычислений, проведенных в примерах 35 и 32). Вычислим последний интеграл:

$$\int_{-\infty}^{+\infty}\sigma\cdot\left(\frac{x-a}{\sigma}\right)^2\cdot\varphi\left(\frac{x-a}{\sigma}\right)dx=\left|\begin{array}{lcc} t=\dfrac{x-a}{\sigma} & \Rightarrow & x=\sigma t+a\Rightarrow dx=\sigma dt\\ x=-\infty & \Rightarrow & t=-\infty\\ x=+\infty & \Rightarrow & t=+\infty\end{array}\right|=\sigma^2\cdot\int_{-\infty}^{+\infty}t^2\varphi(t)dt=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\int_{-\infty}^{+\infty}t^2e^{-\frac{t^2}{2}}dt=$$

$$=\left|\begin{array}{lcr} u=t & \Rightarrow & du=dt\\ dv=te^{-\frac{t^2}{2}}dt & \Rightarrow & v=-e^{-\frac{t^2}{2}}\end{array}\right|=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\left(-te^{-\frac{t^2}{2}}\Bigg|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt\right)=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sigma^2$$

(на последнем шаге мы применили формулу (46)). Итак,

$$M\left(X^2\right)=\sigma^2+a^2. \qquad (56)$$

□

Дадим теперь общее определение независимости (пригодное, в частности, для абсолютно непрерывных случайных величин).

<u>*Определение 24*</u>. Случайные величины *X* и *Y*, определенные на некотором вероятностном пространстве $(\Omega,\mathcal{F},P)$, называются независимыми, если для любых действительных чисел *a, b, c* и *d* выполняется равенство:

$$P\left(a\le X<b,\, c\le Y<d\right)=P(a\le X<b)\cdot P(c\le Y<d). \qquad (57)$$

□

Можно доказать (это выходит за рамки нашей программы), что если с.в. *X* и *Y* независимы, то для любых борелевских множеств $U,V\in\hat{\mathcal{F}}$ справедливо соотношение:

$$P\left(X\in U,\, Y\in V\right)=P(X\in U)\cdot P(Y\in V). \qquad (58)$$

Формула (19) остается справедливой для независимых абсолютно непрерывных с.в. *X* и *Y* (достаточно лишь потребовать существования MX и *MY*). Можно также доказать, что если плотности этих с.в. равны соответственно $p(x)$ и $q(x)$, то с.в. $X\cdot Y$ также имеет плотность $r(x)$, вычисляемую по формуле:

$$r(x)=\int_{-\infty}^{+\infty}p(t)\cdot q\left(\frac{x}{t}\right)\frac{dt}{|t|}. \qquad (59)$$

Дисперсия абсолютно непрерывной случайной величины

Дисперсия абсолютно непрерывной с.в. X определяется формулой (20) (см. определение 15). Используя формулу (53), можно записать:

$$DX = \int_{-\infty}^{+\infty} (x - MX)^2 \cdot p(x) dx \, . \tag{60}$$

Если интеграл справа сходится, то говорят, что дисперсия существует и конечна. Если же он расходится (то есть равен $+\infty$), то говорят, что дисперсия не существует.

Все свойства дисперсии 1–4, доказанные нами в случае конечного вероятностного пространства, остаются справедливыми и для абсолютно непрерывных с.в. При этом только лишь доказательство свойства 2 нуждается в модификации (мы ее опускаем), а доказательства свойств 1, 3 и 4 дословно повторяются (см. раздел "Свойства дисперсии").

Пример 39 (продолжение примеров 30, 33 и 36). Подставив выражения (50) и (54) в формулу (21), вычислим дисперсию с.в. X, равномерно распределенной на отрезке $[a, b]$:

$$DX = M\left(X^2\right) - (MX)^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12} \, . \tag{61}$$

Пример 40 (продолжение примеров 31, 34 и 37). Подставив выражения (51) и (55) в формулу (21), вычислим дисперсию с.в. X, имеющей показательное распределение с параметром λ:

$$DX = M\left(X^2\right) - (MX)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \, . \tag{62}$$

Пример 41 (продолжение примеров 32, 35 и 38). Подставив выражения (52) и (56) в формулу (21), вычислим дисперсию с.в. X, имеющей нормальное распределение с параметрами a и σ:

$$DX = M\left(X^2\right) - (MX)^2 = \sigma^2 + a^2 - a^2 = \sigma^2 \, . \tag{63}$$

Теорема 10. 1) Если с.в. $X \geq 0$ обладает конечным математическим ожиданием, а число $c > 0$, то справедливо ***первое неравенство Чебышева***:

$$P(X \geq c) \leq \frac{MX}{c} \, . \tag{64}$$

2) Если с.в. X имеет конечную дисперсию, а число $c > 0$, то справедливо ***второе неравенство Чебышева***:

$$P(|X - MX| \geq c) \leq \frac{DX}{c^2} \, . \tag{65}$$

Доказательство. 1) Так как $X \geq 0$, то $X \geq X \cdot I_{(X \geq c)} \geq c \cdot I_{(X \geq c)}$. Используя сначала свойство монотонности, а затем линейности математического ожидания, а также свойство 2, позволяющее вычислять математическое ожидание индикатора (см. раздел "Свойства математического ожидания"), получаем:

$$MX \geq M\left(c \cdot I_{(X \geq c)}\right) = c \cdot M\left(I_{(X \geq c)}\right) = c \cdot P(X \geq c) \, ,$$

откуда выводим неравенство (64).

2) Применяя неравенство (64) к случайной величине $(X - MX)^2$, имеем:

$$P(|X - MX| \geq c) = P\left((X - MX)^2 \geq c^2\right) \leq \frac{M\left[(X - MX)^2\right]}{c^2} = \frac{DX}{c^2} \, .$$

□

Лекция 1

Некоторые вероятностные характеристики и распределения

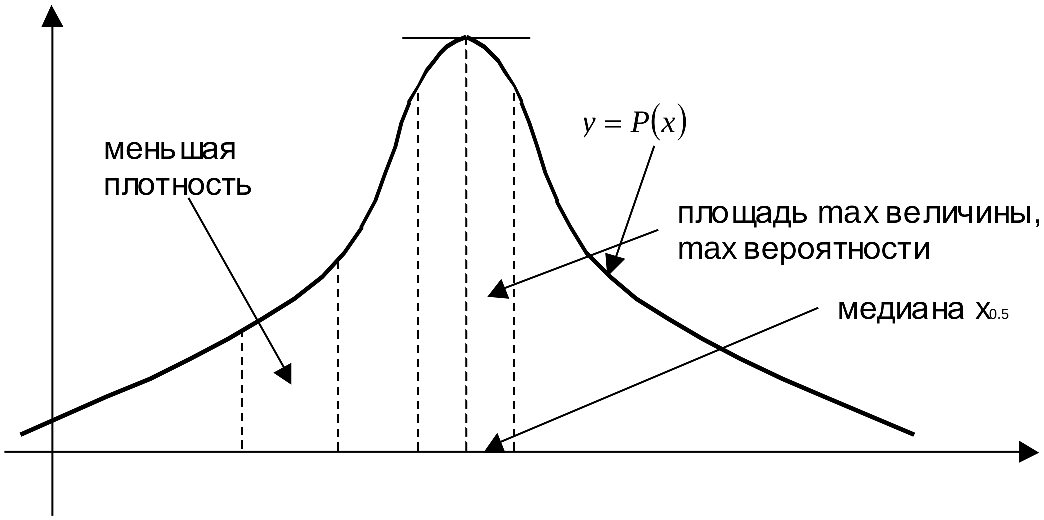
Для простоты будем считать, что все рассматриваемые функции распределения непрерывны, хотя многие определения и результаты верны для произвольных распределений.

Определение 1. ***Модой*** $M_o(X)$ непрерывной случайной величины X называется такое ее значение, при котором достигается максимум плотности распределения.

Определение 2. ***Медианой*** $M_e(X)$ непрерывной случайной величины X называется такое ее значение μ, для которого выполняется равенство

$$P(X<\mu)=P(X>\mu)=0{,}5. \qquad (1)$$

Геометрически мода является абсциссой той точки кривой распределения, ордината которой максимальна. Соответственно, ордината, проведенная в точке с абсциссой $x=\mu$, делит пополам площадь, ограниченную кривой распределения.



Пример 1. Дана плотность вероятности непрерывной случайной величины $p(x)=e^{2x-x^2}$. Найти моду этой случайной величины.

Решение. Найдем максимум функции $y=p(x)$. Для этого находим производные первого и второго порядков: $p'(x)=(2-2x)e^{2x-x^2}$, $p''(x)=-2e^{2x-x^2}+(2-2x)^2e^{2x-x^2}$. Из уравнения $p'(x)=0$ получим $x=1$. Так как $p''(1)=-2<0$, то при $x=1$ функция $y=p(x)$ имеет максимум, то есть $M_o(X)=1$.

Пример 2. Дана плотность вероятности непрерывной случайной величины X

$$p(x)=\begin{cases}0, \ при \ x<0,\\ x-\dfrac{x^3}{4}, \ при \ 0\le x\le 2,\\ 0, \ при \ x>2.\end{cases}$$

Найти медиану этой случайной величины.

<u>*Решение.*</u> Медиану найдем из условия $P(X<\mu)=0{,}5$. По условию $P(X<\mu)=\int\limits_0^{\mu}\left(x-\dfrac{x^3}{4}\right)dx=\left.\left(\dfrac{x^2}{2}-\dfrac{x^4}{16}\right)\right|_0^{\mu}=\dfrac{\mu^2}{2}-\dfrac{\mu^4}{16}$. Таким образом, необходимо решить уравнение $\dfrac{\mu^2}{2}-\dfrac{\mu^4}{16}=0{,}5$ *или* $\mu^4-8\mu^2+8=0$, откуда $\mu^2=4\pm\sqrt{8}$, значит, $\mu=\pm\sqrt{4\pm\sqrt{8}}$. Из полученных корней уравнения нужно выбрать те, которые принадлежат интервалу $(0,2)$. Таким образом, $M_e(X)=\sqrt{4-\sqrt{8}}$.

Математическое ожидание и дисперсия являются частными случаями более общих понятий – моментов случайной величины.

<u>*Определение 3.*</u> ***Начальным моментом порядка*** s случайной величины X называется математическое ожидание s - й степени этой величины:

$$m_s(X)=M\left(X^s\right) \qquad \textbf{(2)}$$

При $s=1$ получаем математическое ожидание случайной величины X.

<u>*Определение 4.*</u> ***Центральным моментом порядка*** s случайной величины X называется математическое ожидание величины $(X-M(X))^s$:

$$\mu_s(X)=M(X-M(X))^s \qquad (3)$$

При $s=2$ получаем дисперсию случайной величины X.

<u>*Теорема 1.*</u> (связь между центральными и начальными моментами).

Для всех $s=0,1,2,\ldots$ справедлива формула

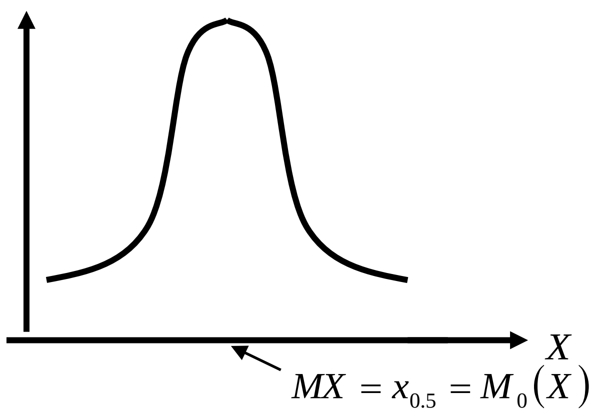
$$\mu_s=\sum_{\kappa=0}^{s}(-1)^k C_s^{\kappa} m_{s-k} m_1^k \qquad (3')$$

.<u>*Доказательство*</u> опускается.

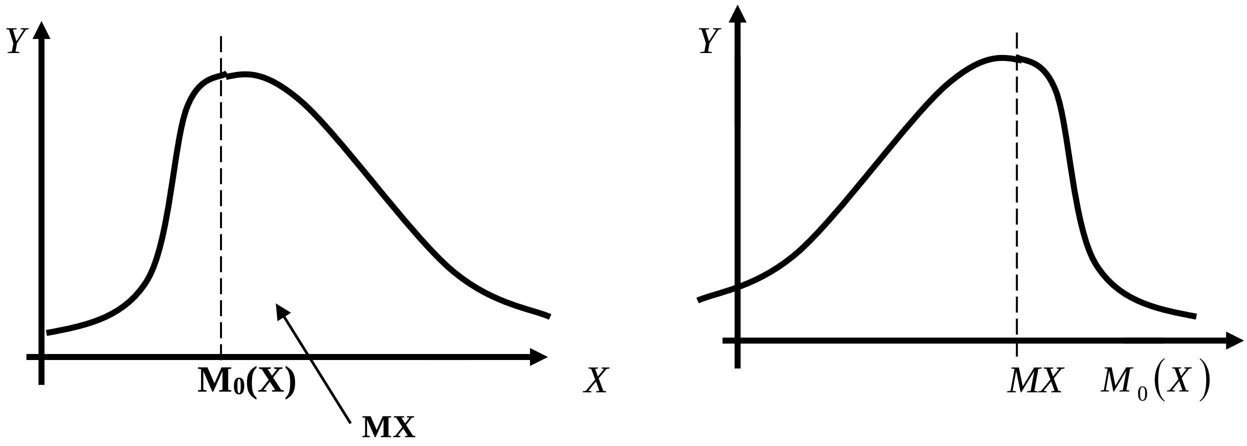
<u>*Определение 5.*</u> ***Коэффициентом асимметрии*** случайной величины X называется число

$$S_k(X)=\frac{\mu_3(X)}{(\sigma(X))^3} \qquad (4)$$

Коэффициент $S_k(X)$ служит для описания отклонения плотностей вероятностей от симметричного вида.



Симметричное распределение



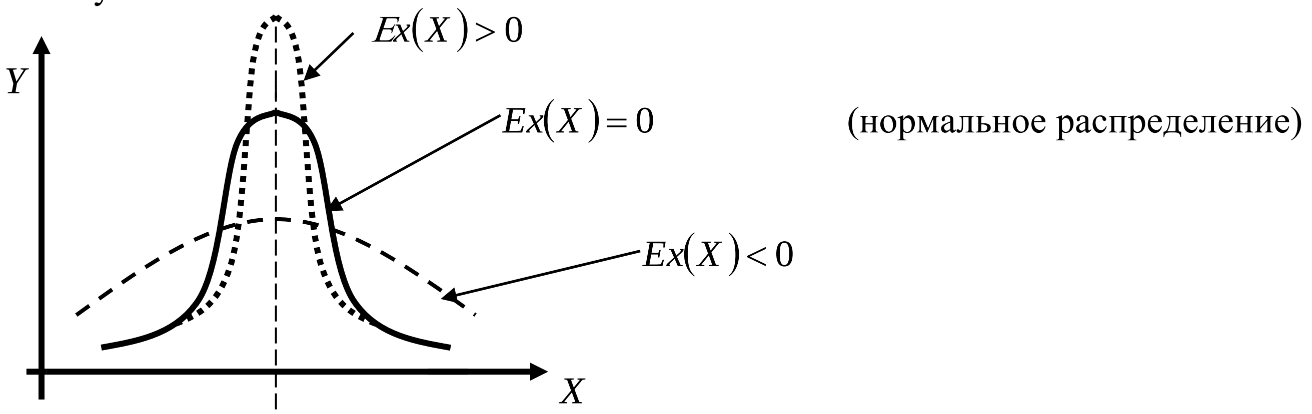
Несимметричное распределение

(интуитивно, большой «хвост» уходит вправо, если $\mu_3(X)>0 \Rightarrow S_k(X)>0$ и соответственно, влево, если $\mu_3(X)<0 \Rightarrow S_k(X)<0$ **).**

<u>*Определение 6.*</u> ***Коэффициентом эксцесса*** случайной величины X называется число

$$Ex(X)=\frac{\mu_4(X)}{(\sigma(X))^4}-3 \qquad (5)$$

Данный коэффициент изучает отклонение от нормальной плотности по части островершинности. При этом "-3" добавлено для того, чтобы для нормального закона распределения $Ex(X)=0$. Положительный эксцесс обычно указывает на то, что рассматриваемое распределение имеет более высокую и более острую вершину, чем у соответствующей нормальной кривой, а отрицательный – более низкую и плоскую.



Существуют также и другие характеристики вероятностных распределений, из которых самыми важными являются производящая функция и характеристическая функция. Последняя функция определяется с помощью конструкций, связанных с теорией функций комплексного переменного.

Производящая функция

С помощью производящих функций удобно находить важнейшие числовые характеристики случайных величин с целыми неотрицательными значениями.

Пусть случайная величина X принимает значения $0,1,2,\ldots k,\ldots$ с вероятностями $p_0, p_1, p_2, \ldots, p_k = P(X=k), \ldots$

Определение 7. Производящей функцией случайной величины *X* называется функция вида

$$\varphi(z)=\sum_{k=0}^{\infty} p_k \cdot z^k = p_0 + p_1 z + p_2 z^2 + \ldots \qquad (6)$$

где z - произвольный параметр, $0 < z \le 1$.

Дифференцируя по z производящую функцию, получим $\varphi'(z)=\sum_{k=0}^{\infty} p_k \cdot k \cdot z^{k-1}$, при $z=1$ $\varphi'(1)=\sum_{k=0}^{\infty} p_k \cdot k = MX = m_1(X)$. Аналогично, взяв вторую производную функции $\varphi(z)$ и положив $z=1$, имеем: $\varphi''(1)=\sum_{k=0}^{\infty} p_k \cdot k^2 - \sum_{k=0}^{\infty} p_k \cdot k = m_2(X) - m_1(X)$. Тогда

$$DX = MX^2 - (MX)^2 = m_2(X) - (m_1(X))^2 = (m_2(X) - m_1(X)) + m_1(X) - (m_1(X))^2 =$$
$$= \varphi''(1) + \varphi'(1) - (\varphi'(1))^2.$$

Пример 3. Найти математическое ожидание и дисперсию случайной величины *X*, заданной рядом распределения:

X	0	1	2	3
p	0,01	0,027	0,243	0,729

Решение. Запишем производящую функцию случайной величины *X:* $\varphi(z)=0{,}01+0{,}027z+0{,}243z^2+0{,}729z^3$. Найдем первую производную этой функции: $\varphi'(z)=0{,}027+0{,}486z+2{,}187z^2$. При $z=1$ получим $\varphi'(1)=2{,}7=MX$. Аналогично находя вторую производную и полагая $z=1$, имеем: $\varphi''(z)=0{,}486+4{,}374z$, $\varphi''(1)=4{,}86$, тогда $DX=4{,}86+2{,}7-(2{,}7)^2=0{,}27$.

Характеристическая функция

Определение 8. Характеристической функцией случайной величины *X* называется математическое ожидание случайной величины e^{itX}:

$$\varphi_X(t) = Me^{itX} \qquad (7)$$

Пусть случайная величина *X* принимает значения 0,1,2,…*k*,... с вероятностями $p_0, p_1, p_2, \ldots, p_k = P(X=k), \ldots$. Тогда ее характеристическая функция определяется формулой: $\varphi_X(t)=\sum_{k=0}^{\infty} e^{itX} \cdot p_k$.

Справедливы следующие равенства:

1. $\varphi_X^{(k)}(0) = i^k \cdot M(X^k) = i^k \cdot m_k(X)$
2. $MX = m_1(X) = -i \cdot \varphi'_X(0)$
3. $DX = m_2(X) - (m_1(X))^2 = -\varphi''_X(0) + (\varphi'_X(0))^2$.

Пример 4. Найти характеристическую функцию, MX, DX случайной величины *X*, распределенной по биномиальному закону *(самостоятельно).*

Лекция 2

Прежде чем перейти к описанию функций распределения, особенно часто встречающейся в статистике, рассмотрим две функции, называемые Г-функциями и β - функциями Эйлера.

<u>*Определение 9.*</u> ***Гамма-функциями Эйлера*** (Г-функциями Эйлера) называется функция вида**:**

$$\Gamma(n)=\int_0^\infty t^{n-1}e^{-t}\,dt \qquad (8)$$

где $n>0$ - любое положительное число.

Например, $\Gamma(1)=\int_0^\infty t^0e^{-t}dt=\int_0^\infty e^{-t}dt=1$**.**

<u>*Теорема 2.*</u> Справедливы следующие равенства:

1) $\Gamma(n+1)=n\Gamma(n)$, для $\forall n>0$;

2) $\Gamma(n+1)=n!, n=1,2,3...$

<u>*Доказательство:*</u>

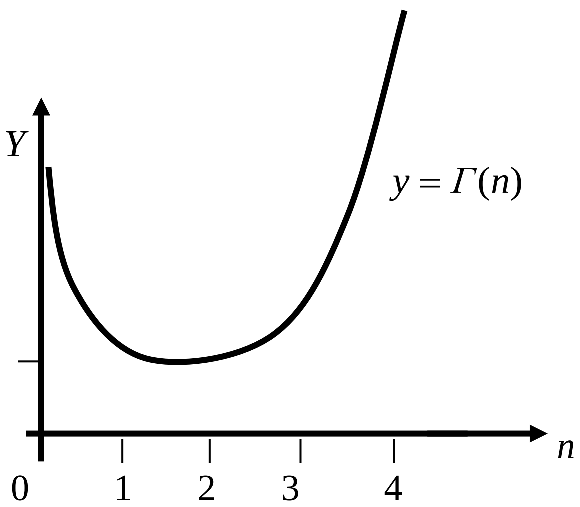
1) Имеем:

$$\Gamma(n+1)=\int_0^\infty t^ne^{-t}dt=\left|\begin{matrix} n=t^n \Rightarrow dn=nt^{n-1}dt \\ dv=e^{-t}dt \Rightarrow v=-e^{-t}\end{matrix}\right|=-t^ne^{-t}\Big|_0^\infty+n\int_0^\infty t^{n-1}e^{-t}dt=n\Gamma(n).$$

2) В силу доказанного выше

$$\Gamma(n+1)=n\Gamma(n)=n(n-1)\Gamma(n-1)=n(n-1)(n-2)\Gamma(n-2)...=n(n-1)(n2)...1\Gamma(1)=n!$$

Данная теорема показывает, что Г-функция обобщает понятие факториала.



<u>*Упражнение.*</u> Доказать , что $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$.

Введем следующие обозначения: $\Gamma(x,n)=\int_0^x t^{n-1}e^{-t}dt$ - неполная Γ - функция,

$I(x,n)=\dfrac{\Gamma(x,n)}{\Gamma(n)}$ -отношение неполной Γ - функции к полной Γ - функции.

<u>*Определение 10.*</u> ***Бета - функцией Эйлера*** **(** B **-функция Эйлера)** называется функция двух переменных вида

$$B(u,v)=\int_0^1 t^{u-1}(1-t)^{v-1}(1-t)^{v-1}dt \qquad (9)$$

где все $u,v>0$.

<u>*Теорема 3.*</u> (о связи Γ - функций и B- функций). Справедливы следующие равенства: 1) $B(u,v)=\dfrac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$ для $\forall\ u,v>0$.

2) $B(u,v)=B(v,u)$.

Будем употреблять следующие обозначения:

$B_x(u,v)=\int_0^x t^{x-1}(1-t)^{v-1}dt$–неполная B - функция,

$I_{x(u,v)}=\dfrac{B_x(u,v)}{B(u,v)}$-отношение неполной B - функции к полной B- функции.

<u>*Теорема 4*</u> . Справедливо следующее равенство $I_x(u,v)=I_{1-x}(v,u)$**.**

Распределение χ^2 (хи – квадрат или Пирсона)

<u>*Определение 11.*</u> Если $u_{1,}u_2,u_3,...,u_n$ - независимые с.в., каждая из которых имеет нормальное распределение $N(0,1)$ то распределение с.в. $u_1^2+u_2^2+...+u_n^2$ называется χ^2 распределением с n- степенями свободы.

Плотность этого распределения имеет вид: $p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\cdot\Gamma\left(\dfrac{n}{2}\right)}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}},\ при\ x>0,\\ 0,\ при\ x\le 0.\end{cases}$

<u>*Упражнение.*</u> Доказать, что функция $p(x)$ является плотностью некоторого вероятностного распределения.

Функция распределения: $F(x)=I(\dfrac{x}{2},\dfrac{n}{2})$

Математическое ожидание: $M(X)=n$. ***Дисперсия:*** $D(X)=2n$.

Медиана: $x_{0.5}\approx n-0{,}67$.

Мода: $M_0(X)=n-2$ при $n\ge 2$**,** при $n=1$ $M_0(X)=0$**.**

Коэффициент асимметрии: $S_k=\sqrt{\dfrac{8}{n}}$. ***Коэффициент эксцесса:*** $E_x=\dfrac{12}{n}$.

Начальные моменты: $m_s=\prod_{k=0}^{s-1}(n+2k)$.

Центральные моменты: $\mu_3=8n,\ \mu_4=12n(n+4)$.

Распределение Стьюдента

<u>*Определение 12.*</u> Пусть Z - нормально распределенная случайная величина $N(0,1)$, а V – независимая от Z величина, которая распределена по закону χ^2 с n степенями свободы. Тогда величина $\mathrm{T} = \dfrac{Z}{\sqrt{V/n}}$ имеет распределение, называемое распределением Стьюдента с n степенями свободы.

Плотность этого распределения имеет вид:

$$p(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (n \text{ - степень свободы}).$$

Функция распределения хорошо не вычисляется.

***Математическое ожидание*:** $MX = 0$ *при* $n > 1$.

Дисперсия: $DX = \dfrac{n}{n-2}$ *при* $n > 2$.

Медиана: $x_{0,5} = 0$.

Коэффициент асимметрии: $S_k = 0$, *при* $n > 3$.

Коэффициент эксцесса: $Ex(X) = \dfrac{6}{n-4}$, *при* $n > 4$.

Начальные и центральные моменты совпадают (т.к. $MX = 0$), поэтому:

$$m_s = \mu_s = \begin{cases} 0, & \text{если } s - \text{нечетный порядок} \\ \dfrac{1\cdot 3\cdot 5 \ldots \cdot (s-1)}{(n-2)(n-4)\ldots(n-s)}\, n^{\frac{s}{2}}, & \text{если } s - \text{четный порядок} \end{cases}.$$

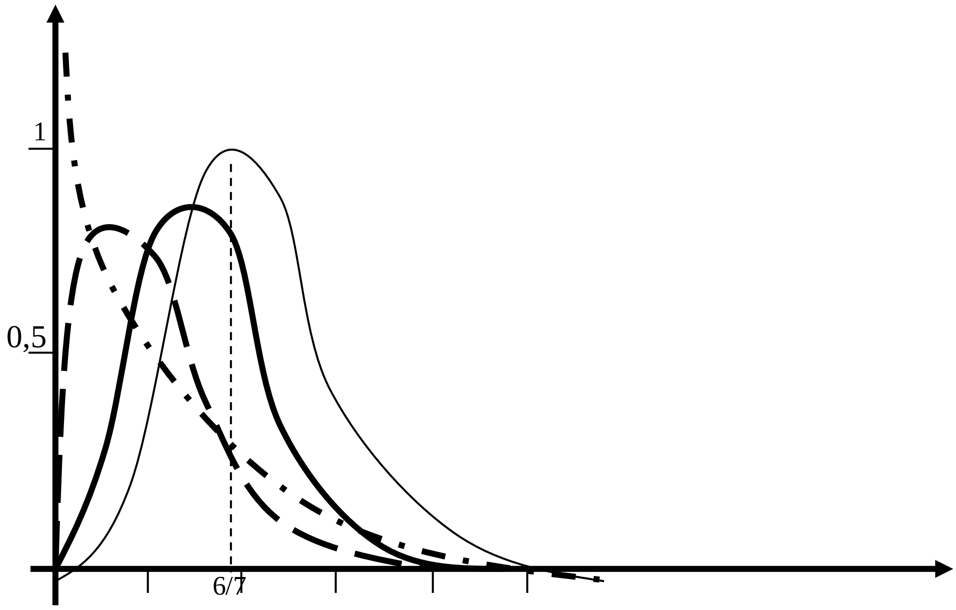
Распределение Фишера – Снедекора

<u>*Определение 13.*</u> Пусть U и V – независимые случайные величины, распределенные по закону χ^2 со степенями свободы u и v соответственно. Тогда величина $F = \dfrac{U/u}{V/v} = \dfrac{U\cdot v}{V\cdot u}$ имеет распределение, называемое распределением Фишера – Снедекора со степенями свободы u и v.

Плотность этого распределения имеет вид:

$$p(x) = \begin{cases} \dfrac{1}{B\left(\frac{u}{2}, \frac{v}{2}\right)}\left(\dfrac{u}{v}\right)^{\frac{u}{2}} x^{\frac{u}{2}-1}\left(1+\dfrac{u}{v}x\right)^{-\frac{u+v}{2}}, & \text{при } x > 0 \\ 0, & \text{если } x \le 0 \end{cases}$$

График плотностей:



Функция распределения: $F(x) = I_{\frac{u+x}{v+ux}}\left(\frac{u}{2}, \frac{v}{2}\right)$.

***Математическое ожидание*:** $MX = \frac{v}{v-2}$, *если* $v > 2$.

***Мода*:** $M_0(X) = \frac{v(x-2)}{u(v+2)}$, *если* $u \geq 2$.

Дисперсия: $D(X) = \frac{2v^2(u+v-2)}{u(v-2)^2(v-4)}$, *если* $v > 4$.

Коэффициент асимметрии: $S_k(X) = \frac{2(2u+v-1)}{v-6}\sqrt{\frac{2(v-4)}{u(u+v-2)}}$, *если* $v > 0$

***Коэффициент эксцесса*:** $E_x(X) = \frac{12\left[(v-2)^2(v-4) + u(5v-22)(u+v-2)\right]}{u(u+v-2)(v-6)(v-8)}$, *если* $v > 8$

Начальные моменты: $m_s = (\frac{u}{v})^2 \frac{u(u+2)(u+6)...(u+2s-1)}{(v-2)(v-4)(v-6)...(v-2s)}$, *если* $v > 2s$.

Лекция 3

Ковариация двух случайных величин. Коэффициент корреляции

Рассмотрим случайные величины ξ_1, ξ_2 . Заметим, что если ξ_1=const, то ξ_1, ξ_2 — независимы. Рассмотрим случай, когда ξ_1, ξ_2 — индикаторы: $\xi_1=I_A$, $\xi_2=I_B$.

Можно считать, что если события, связанные с принятием данными случайными величинами каких-то значений, независимы, то и случайные величины будут независимыми.

Введем следующие обозначения:

$\{\xi_1=1\}=A$, $\{\xi_2=1\}=B$, $\{\xi_1=0\}=\overline{A}$, $\{\xi_2=0\}=\overline{B}$, тогда

$$P(\{\xi_1=1\}\cdot\{\xi_2=1\})=P(AB)=P(A)\cdot P(B),$$
$$P(\{\xi_1=1\}\cdot\{\xi_2=0\})=P(A\overline{B})=P(A)\cdot P(\overline{B}),$$
$$P(\{\xi_1=0\}\cdot\{\xi_2=1\})=P(\overline{A}B)=P(\overline{A})\cdot P(B),$$
$$P(\{\xi_1=0\}\cdot\{\xi_2=0\})=P(\overline{A}\cdot\overline{B})=P(\overline{A})\cdot P(\overline{B}).$$

Если все это выполняется, то случайные величины ξ_1 и ξ_2 — независимы.

Легко доказать, что в случае индикатора нужно требовать только выполнение равенства $P(AB)=P(A)\cdot P(B)$, а остальные соотношения можно вывести из него:

$$P(A\cdot\overline{B})=P(A\cdot(\omega-B))=P(A\omega-AB)=P(A\omega)-P(AB)=P(A)-P(A)\cdot P(B)$$
$$P(A)\cdot P(\overline{B})=P(A)\cdot(1-P(B))=P(A)-P(AB)=P(A)\cdot(P(B)+P(\overline{B}))-P(AB)=P(A)\cdot$$
$$\cdot P(B)+P(A)\cdot P(\overline{B})-P(AB)=P(AB)+P(A\overline{B})-P(AB)=P(A\overline{B}).$$

<u>*Определение 14*</u>. Дискретные случайные величины ξ_1 и ξ_2 называются ***независимыми,*** если $P(\{\xi_1=x_i\}\cdot\{\overline{\xi}_2=y_j\})=P(\xi_1=x_i)\cdot P\{\xi_2=y_j\}$, где x_1, x_2, …— множество значений случайной величины ξ_1, а y_1, y_2, … — множество значений случайной величины ξ_2.

<u>*Определение 15*</u>. Если ξ_1, ξ_2 — произвольные случайные величины, то их ***независимость*** определяется следующим образом:

$$P(\{\xi_1<x\}\cdot\{\xi_2<y\})=P(\xi_1<x)\cdot P(\xi_2<y),\ \forall x,y\in R$$

Можно доказать, что при выполнении данного соотношения выполняется и следующее соотношение:

$$P(\{\xi_1\in B_1\}\cdot\{\xi_2\in B_2\})=P(\xi_1\in B_1)\cdot P(\xi_2\in B_2),$$

где B_1, B_2 — произвольные подмножества R.

В нашем определении: $B_1=(-\infty;x)$, $B_2=(-\infty;y)$.

<u>*Определение 16*</u>. ***Ковариацией*** случайных величин ξ_1 и ξ_2 называется число

$$\mathrm{cov}(\xi_1,\xi_2)=M[(\xi_1-M\xi_1)\cdot(\xi_2-M\xi_2)] \qquad (10)$$

Свойства ковариации.

1. Если $\xi_1=\xi_2$, то $\operatorname{cov}(\xi_1,\xi_2)=D\xi_1$.

2. Если случайные величины ξ_1 и ξ_2 независимы, то их ковариация равна нулю: $\operatorname{cov}(\xi_1,\xi_2)=0$

Доказательство. Так как ξ_1 и ξ_2 независимы, а $M\xi_1$ и $M\xi_2$ постоянны, то $\xi_1-M\xi_1$ и $\xi_2-M\xi_2$ тоже независимы. Тогда

$$\operatorname{cov}(\xi_1,\xi_2)=M(\xi_1-M\xi_1)\cdot M(\xi_2-M\xi_2)=(M\xi_1-M\xi_1)\cdot M(\xi_2-M\xi_2)=0.$$

3. Справедлива формула: $\operatorname{cov}(\xi_1,\xi_2)=M(\xi_1,\xi_2)-M\xi_1\cdot M\xi_2$

Доказательство.

$$\operatorname{cov}(\xi_1,\xi_2)=M[\xi_1\cdot\xi_2-\xi_1 M\xi_2-\xi_2 M\xi_1+M\xi_1\cdot M\xi_2]=M(\xi_1,\xi_2)-M\xi_2\cdot M\xi_1-$$
$$-M\xi_1\cdot M\xi_2+M\xi_1\cdot M\xi_2=M(\xi_1,\xi_2)-M\xi_1\cdot M\xi_2.$$

4. $\operatorname{cov}(\xi_1,\xi_2)=\operatorname{cov}(\xi_2,\xi_1)$.

5. Ковариация линейна отдельно по каждой своей переменной:

$$\operatorname{cov}(\alpha\xi_1+\beta\xi_2,\xi_3)=\alpha\operatorname{cov}(\xi_1,\xi_3)+\beta\operatorname{cov}(\xi_2,\xi_3)$$

(доказательство самостоятельно).

Определение 17. Пусть $\xi_1, \xi_2,.., \xi_n$ — набор случайных величин, у которых существуют попарные ковариации, т.е. $\sigma_{i,j}=\operatorname{cov}(\xi_i,\xi_j),\ i,j=1,..,n$. Матрица

$$\sigma=\begin{vmatrix}\sigma_{11} & \sigma_{12} & \dots & \sigma_{1n}\\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n}\\ \dots & \dots & \dots & \dots\\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn}\end{vmatrix} \qquad (11)$$

называется ***ковариационной матрицей*** случайных величин $\xi_1, \xi_2,.., \xi_n$.

Теорема 5. Пусть случайные величины $\xi_1, \xi_2,.., \xi_n$ имеют ковариационную матрицу σ, а $c_1, c_2,\ldots c_n$ — действительные числа, тогда

$$\operatorname{cov}\left(\sum_{i=1}^n c_i\xi_i,\sum_{j=1}^n c_j\xi_j\right)=D\left(\sum_{i=1}^n c_i\xi_i\right)=\sum_{i,j=1}^n c_i c_j\sigma_{i,j}$$

Доказательство.

$$D\left(\sum_{i=1}^n c_i\xi_i\right)=M\left[\sum_{i=1}^n c_i\xi_i-M\left(\sum_{i=1}^n c_i\xi_i\right)\right]^2=M\left[\sum_{i=1}^n c_i\xi_i-\sum_{i=1}^n c_i M\xi_i\right]^2=M\left[\sum_{i=1}^n c_i(\xi_i-M\xi_i)\right]^2=$$

$$= M\left[\left(\sum_{i=1}^{n} c_i\left(\xi_i - M\xi_i\right)\right)\cdot\left(\sum_{j=1}^{n} c_j\left(\xi_j - M\xi_j\right)\right)\right] = M\left[\sum_{i,j=1}^{n} c_i c_j\left(\xi_i - M\xi_i\right)\cdot\left(\xi_j - M\xi_j\right)\right] =$$

$$= \sum_{i,j=1}^{n} c_i c_j \cdot M\left[\left(\xi_i - M\xi_i\right)\cdot\left(\xi_j - M\xi_j\right)\right] = \sum_{i,j=1}^{n} c_i c_j \sigma_{i,j}$$

Следствие. $\left|\operatorname{cov}\left(\xi_1,\xi_2\right)\right| = \sqrt{D\xi_1}\cdot\sqrt{D\xi_2}$

Доказательство. В условии теоремы 5 положим, что n=2, c_1=x, c_2=1. Имеем:

$$D\left(x\xi_1+\xi_2\right) = \sum_{i,j=1}^{2} c_i c_j \sigma_{i,j} = c_1c_1\sigma_{1,1} + c_1c_2\sigma_{1,2} + c_2c_1\sigma_{2,1} + c_2c_2\sigma_{2,2} = \sigma_{1,1}x^2 + 2\sigma_{1,2}x + \sigma_{2,2}.$$

Так как $D\left(x\xi_1+\xi_2\right) \ge 0$, то для всех действительных $x \in R$ $\sigma_{1,1}x^2 + 2\sigma_{1,2}x + \sigma_{2,2} \ge 0 \Rightarrow$

$$D\left(\sigma_{1,1}x^2 + 2\sigma_{1,2}x + \sigma_{2,2}\right) \le 0 \Rightarrow \sigma_{1,2}^2 \le \sigma_{1,1}\sigma_{2,2} \Rightarrow \left|\sigma_{1,2}^2\right| \le \sqrt{\sigma_{1,1}}\cdot\sqrt{\sigma_{2,2}}.$$

Определение 17. ***Коэффициентом корреляции*** случайных величин ξ_1 и ξ_2 называется число

$$cor\left(\xi_1,\xi_2\right) = \frac{cov\left(\xi_1,\xi_2\right)}{\sqrt{D\xi_1}\cdot\sqrt{D\xi_2}} \tag{12}$$

Коэффициент корреляции служит для количественного изучения степени зависимости случайных величин ξ_1 и ξ_2.

Свойства коэффициента корреляции.

1. Если случайные величины ξ_1 и ξ_2 независимы, то $cor\left(\xi_1,\xi_2\right) = 0$.

2. $\left|cor\left(\xi_1,\xi_2\right)\right| \le 1$, причем $cor\left(\xi_1,\xi_2\right) = 1$ тогда и только тогда, когда между случайными величинами ξ_1 и ξ_2 существует линейная зависимость, т.е. либо $\xi_2 = a\xi_1 + b$, либо $\xi_1 = a\xi_2 + b$.

Доказательство. Докажем только одну сторону этого утверждения: если $\xi_2 = a\xi_1 + b$, то $\left|cor\left(\xi_1,\xi_2\right)\right| = 1$.

$$D\xi_2 = D\left(a\xi_1\right) + Db = a^2 D\xi_1$$

$$\operatorname{cov}\left(\xi_1,\xi_2\right) = \operatorname{cov}\left(\xi_1, a\xi_1 + b\right) = a\operatorname{cov}\left(\xi_1,\xi_1\right) + b\operatorname{cov}\left(\xi_1, 1\right) = aD\xi_1$$

$$\left|cor\left(\xi_1,\xi_2\right)\right| = \frac{|a| D\xi_1}{\sqrt{D\xi_1}\cdot\sqrt{a^2 D\xi_1}} = 1.$$

3. Из того, что $cor\left(\xi_1,\xi_2\right) = 0$, вообще говоря, **не** следует, что случайные величины ξ_1 и ξ_2 независимы. *(Доказательство самостоятельно)*

<u>Пример 5</u>. В двух ящиках находятся по шесть шаров. В первом ящике – 1 шар с номером 1, 2 шара с номером 2, 3 шара с номером 3; во втором ящике – 2 шара с номером 1, 3 шара с номером 2, 1 шар с номером 3. Пусть X – номер шара, вынутого из первого ящика, Y – номер шара, вынутого из второго ящика. Из каждого ящика вынули по шару. Найти коэффициент корреляции.

<u>Решение.</u>

Пусть Y=1, тогда по условию задачи получаем, что значение X=1 повторяется 3 раза, значение X=2 повторяется 4 раза, значение X=3 повторяется 6 раз. Значит, при Y=1 получаем ряд распределения случайной величины X:

x_i	1	2	3
p_i	$1/6$	$1/3$	$1/2$

Если Y=2, то значение X=1 повторяется 3 раза, значение X=2 повторяется 6 раз, значение X=3 повторяется 9 раз. Следовательно, при Y=2 получается ряд распределения случайной величины X:

x_i	1	2	3
p_i	$1/6$	$1/3$	$1/2$

При Y=3 значение X=1 повторяется 1 раз, значение X=2 – 2 раза, значение X=3 повторяется 3 раза. Поэтому при Y=3 получается следующий ряд распределения случайной величины X:

x_i	1	2	3
p_i	$1/6$	$1/3$	$1/2$

Таким образом, при различных значениях Y получаем один и тот же ряд распределения случайной величины X. Так как ряд распределения случайной величины X не зависит от значений случайной величины Y, то случайные величины X и Y независимы. Тогда по свойству 1 коэффициент корреляции равен нулю.

Лекция 4

Обработка результатов измерения

Пусть мы производим измерение. Например, контролируем размеры производимой на заводе детали или делаем опрос общественного мнения. В результате получаем набор чисел $x_1, x_2, \ldots, x_n$ — ***выборку измерений с объемом выборки n***.

Если повторить эти измерения вновь и оставить объем выборки прежним, то получим новый набор чисел. Выборки производятся для того, чтобы понять свойства объекта исследования. Об объекте исследования можно мыслить как о некоторой случайной величине X. Таким образом, производя измерения, мы получаем в результате значение этой случайной величины: $x_1, x_2, \ldots, x_n$. Снова произведя измерения, мы получим новые значения случайной величины: $x_1, x_2, \ldots, x_n$.

Самый естественный способ проведения измерений:

1) считать, что все измерения произведены независимо друг от друга;

2) следует считать, что в результате измерений мы получаем случайные величины, имеющие ту же функцию распределения, что и измеряемая величина X.

Таким образом, мы получаем в результате измерений набор случайных величин $X_1, X_2, \ldots, X_n$.

Определение 18. ***Генеральной совокупностью*** называется совокупность всех подлежащих изучению объектов или возможных результатов всех мыслимых наблюдений, производимых в неизменных условиях над одним объектом.

□

Определение 19. ***Выборкой*** объема ***n***, измеряющего случайную величину X, называется набор случайных величин $X_1, X_2, \ldots, X_n$, удовлетворяющих следующим условиям:

1) случайные величины $X_1, X_2, \ldots, X_n$ независимы в совокупности;

2) $P(X_k < x) = P(X < x)$, $\forall x \in R$, $\forall k = 1,2,\ldots,n$.

□

Это равенство означает совпадение функций распределения случайных величин X_k и X.

При составлении выборки после того, как объект отобран и над ним произведено наблюдение, он может быть возвращен в генеральную совокупность или не возвращен. В связи с этим различают повторные и бесповторные выборки.

Определение 20. ***Повторной*** называется выборка, при которой отобранный объект возвращается в генеральную совокупность перед извлечением следующего объекта.

□

Определение 21. ***Бесповторной*** называется выборка, при которой отобранный объект не возвращается в генеральную совокупность.

□

Определение 22. ***Реализацией выборки*** называют конкретные значения $x_1, x_2, ..., x_n$ выборки, полученные в результате наблюдений (испытаний).

Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была ***репрезентативной***, то есть наиболее полно представляла изучаемые признаки генеральной совокупности. Согласно закону больших чисел условием обеспечения репрезентативности выборки является соблюдение случайности отбора, то есть все объекты генеральной совокупности должны иметь равные вероятности попасть в выборку.

Пример 6. Двадцать студентов проходят тестирование по математике. Каждый из них может набрать от 0 до 5 баллов. Пусть X_k – количество баллов, полученное k- ым студентом, $k = \overline{1,20}$. Тогда значения 0,1,2,3,4,5 – все возможные количества баллов, набранные одним студентом – образуют генеральную совокупность. Результат тестирования двадцати студентов является выборкой $X_1, X_2, ..., X_{20}$. Реализациями выборки могут быть следующие наборы чисел: {0, 1, 5, 3, 4, 2, 3, 1, 4, 4, 3, 5, 5, 2, 1, 0, 1, 1, 2, 4} или {3, 1, 1, 0, 4, 5, 3, 5, 4, 5, 3, 2, 5, 1, 1, 4, 2, 3, 5, 0}.

□

Распределение выборки. Эмпирическая функция распределения

Предположим, что мы изучаем некоторую случайную величину X и с этой целью производим ряд независимых наблюдений. Пусть X приняла n_1 раз значение, равное x_1, n_2 раз – значение x_2, …, n_k раз – значение x_k, при этом $n_1 + n_2 + ... + n_k = n$, где n – объем выборки.

Определение 23. ***Вариантами*** случайной величины X называются значения x_k.

□

Определение 24. ***Частотами*** называются числа n_k, которые показывают, сколько раз встречаются значения x_k в ряде наблюдений.

□

Определение 25. ***Относительными частотами*** называются отношение частот n_k к объему выборки n:

$$\hat{p}_k = \frac{n_k}{n} \qquad (13)$$

□

Определение 26. ***Распределением выборки или вариационным рядом*** называется ранжированный (упорядоченный) перечень вариантов x_k и соответствующих им частот n_k. Вариационный ряд называется ***дискретным***, если любые его варианты отличаются на постоянную величину, и ***интервальным***, если варианты могут отличаться один от другого на сколь угодно малую величину.

□

Для построения интервального ряда необходимо определить величину частичных интервалов, на которые разбивается весь интервал варьирования наблюдаемых значений случайной величины X. Длину частичного интервала h нужно выбрать таким образом, чтобы построенный ряд, с одной стороны, не был громоздким, а с другой стороны, позволял выявить характерные черты изменения

изучаемой случайной величины X. По ***формуле Стерджеса*** оптимальное число интервалов определяется по формуле

$$m = 1 + 3{,}322 \cdot \lg n \qquad (14)$$

а длина интервала –

$$h = \frac{x_{\max} - x_{\min}}{m} \qquad (15)$$

где $x_{\max} - x_{\min}$ – разность между наибольшим и наименьшим наблюдаемыми значениями, при этом за начало первого интервала принимается

$$x_{нач} = x_{\min} - 0{,}5 \cdot h$$

Пример 7. В результате трех экзаменов группа из 30 наудачу выбранных абитуриентов набрала следующую сумму баллов: 157, 175, 170, 166, 159, 173, 182, 167, 171, 169, 172, 164, 173, 175, 171, 158, 179, 156, 165, 179, 155, 178, 160, 154, 183, 153, 155, 167, 186, 163. Построить интервальный ряд.

Решение. Сначала упорядочим полученные данные по возрастанию:
153, 154, 155, 155, 156, 157, 158, 159, 160, 163, 164, 165, 166, 167, 167, 169, 170, 171, 171, 172, 173, 173, 175, 175, 178, 179, 179, 182, 183, 186.
Ясно, что $x_{\min} = 153$, $x_{\max} = 186$.
Найдем число частичных интервалов и длину интервала по формуле Стерджеса: $m = 1 + 3{,}322 \cdot \lg 30 \approx 5{,}91$, $h = \frac{186 - 153}{5{,}91} \approx \frac{33}{5{,}91} \approx 5{,}59$. Возьмем $h = 6$, тогда $x_{нач} = 153 - 0{,}5 \cdot 6 = 150$. Разобъем весь ряд на 6 интервалов: [150, 156), [156, 162), [162, 168), [168, 174), [174, 180), [180, 186). Подсчитаем число абитуриентов, попавших в каждый из полученных интервалов, и получим интервальный ряд:

Сумма баллов	[150, 156)	[156, 162)	[162, 168)	[168, 174)	[174, 180)	[180, 186)
Частота	4	5	6	7	5	3
Относит. частота	$4/30$	$5/30$	$6/30$	$7/30$	$5/30$	$3/30$

□

Определение 27. ***Эмпирической функцией распределения*** называется функция $\hat{F}_n(x)$, определяющая для каждого значения x относительную частоту события $X < x$:

$$\hat{F}_n(x) = \hat{p}_n(X < x) \qquad (16)$$

□

Обозначим через n_x число наблюдений, меньших x, тогда значения эмпирической функции распределения можно находить по формуле:

$$\hat{F}_n(x) = \frac{n_x}{n} \qquad (17)$$

В отличие от эмпирической функции распределения $\hat{F}_n(x)$ функцию распределения $F(x)$ генеральной совокупности называют ***теоретической функцией распределения***. Главное различие между эмпирической и теоретической

функциями распределения состоит в том, что теоретическая функция $F(x)$ определяет вероятность события $X < x$, а эмпирическая функция $\hat{F}_n(x)$ – относительную частоту этого события.

Полигон и гистограмма

Наблюдаемые данные, представленные в виде вариационного ряда, можно изобразить графически с помощью полигона и гистограммы. Это позволяет получить наглядное представление о закономерности варьирования наблюдаемых значений случайной величины X.

Определение 28. ***Полигоном частот*** называется ломаная, отрезки которой соединяют точки с координатами (x_1, n_1), (x_2, n_2), …, (x_k, n_k); ***полигоном относительных частот*** – с координатами $(x_1, \hat{p}_1)$, $(x_2, \hat{p}_2)$, …, $(x_k, \hat{p}_k)$.

□

Определение 29. ***Гистограммой частот (относительных частот)*** называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых являются частичные интервалы длиной h, а высоты равны частотам (относительным частотам) соответствующих интервалов.

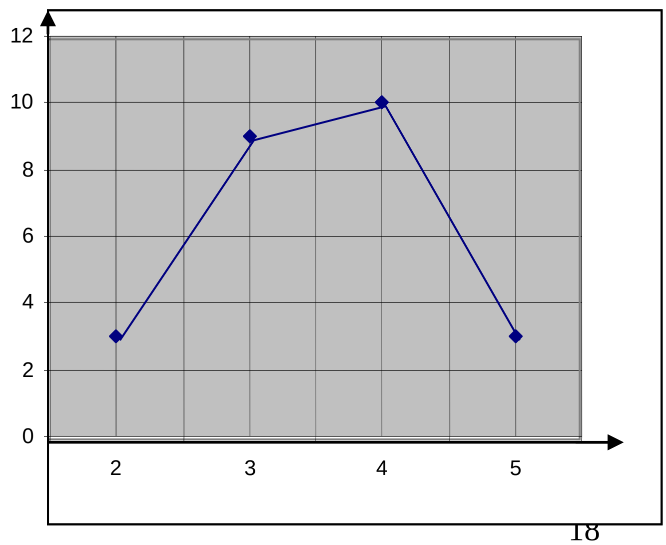
□

Пример 8. Для оценивания знаний студентов-первокурсников проведена контрольная работа по высшей математике. Результаты контроля в выбранной группе из 25 студентов оказались следующими: 3 студента выполнили работу на «5», 10 студентов – на «4», 9 студентов – на «3» и 3 студента – на «2». Построить полигон частот.

Решение. Представим исходные данные в виде дискретного вариационного ряда:

x_k	2	3	4	5
n_k	3	9	10	3

Построим полигон частот:



Лекция 5

Точечные оценки параметров распределения

Будем считать, что измеряемая случайная величина X имеет неизвестные параметры, которые нам нужно оценить. Например, мы можем знать, что случайная величина X имеет нормальное распределение $N(a,\sigma^2)$, но параметры a и σ нам неизвестны.

Для того чтобы интуитивно понять смысл дальнейших вычислений, вернемся к исходному набору чисел $x_1, x_2, ..., x_n$. Ввиду предположений о том, как проводятся наши измерения, можем сделать вывод, что числа $x_1, x_2, ..., x_n$ появляются равновероятно. Таким образом, можно записать следующий закон распределения:

	1	2	..	n
	/n	/n	..	/n

Определение 30. ***Выборочным средним*** $\overline{X}$ (средним арифметическим) наблюдаемых значений случайной величины X называется число, определяемое формулой:

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k \qquad (18)$$

Если наблюдаемые данные представлены в виде дискретного ряда, где $x_1, x_2, ..., x_k$ - варианты значений случайной величины X, а $n_1, n_2, ..., n_k$ - соответствующие им частоты, то выборочное среднее вычисляется по формуле

$$\overline{X} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + ... + x_n \cdot n_k}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k \cdot n_k = \sum_{k=1}^{n} x_k \cdot \hat{p}_k \qquad (19)$$

Определение 31. ***Выборочной дисперсией*** $\hat{D}X$ значений случайной величины X называется среднее арифметическое квадратов отклонений наблюдаемых значений этой величины от их выборочного среднего:

$$\hat{D}X = \frac{1}{n}\sum_{k=1}^{n}\left(x_k - \overline{X}\right)^2 \qquad (20)$$

Аналогично для дискретного вариационного ряда выборочная дисперсия определяется формулой:

$$\hat{D}X = \frac{1}{n}\sum_{k=1}^{n}\left(x_k - \overline{X}\right)^2 \cdot n_k = \sum_{k=1}^{n}\left(x_k - \overline{X}\right)^2 \cdot \hat{p}_k \qquad (21)$$

Интуиция нам подсказывает, что числа $\overline{X}$ и $\hat{D}X$ должны быть приближениями математического ожидания и дисперсии случайной величины X. Оказывается, что первая формула — это хорошее приближение математического ожидания случайной величины X, а вторая формула — не очень хорошее приближение дисперсии случайной величины X. Поэтому вводится следующая ***исправленная дисперсия***:

$$\hat{S}^2 = \frac{n}{n-1}\hat{D}X = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \overline{X}\right)^2 \qquad (22)$$

Данное выражение будет давать хорошее приближение дисперсии случайной величины X.

Определение 32. Выборочным средним квадратическим отклонением $\hat{\sigma}_x$ называется арифметический квадратный корень из выборочной дисперсии:

$$\hat{\sigma}_x = \sqrt{\hat{D}X} \qquad (23)$$

Определение 33. Пусть закон распределения случайной величины X содержит неизвестный параметр θ. ***Оценкой*** параметра θ называется некоторая функция $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, ..., X_n)$ от случайных величин $X_1, X_2, ..., X_n$.

□

Определение 34. Оценка $\hat{\theta}_n$ называется ***несмещенной***, если $M\hat{\theta}_n = \theta$.

Определение 35. Оценка $\hat{\theta}_n$ называется ***состоятельной***, если для всякого $\varepsilon > 0$ $\lim\limits_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) \le 1$.

□

В теории вероятности в этом случае говорят, что $\hat{\theta}_n \to \theta$ (по вероятности).

Определение 36. Оценка $\hat{\theta}_n$ называется ***эффективной***, если для любой другой оценки θ_n' параметра θ выполняется соотношение $D\hat{\theta}_n \le D\theta_n'$.

Несмещенность оценки означает, что прибор, которым мы производили измерения, либо способ измерения не содержит системной ошибки. В среднем мы получаем измеряемый параметр θ. Состоятельность ошибки говорит о том, что при увеличении числа измерений наша оценка приближается к измеряемому параметру θ. А эффективность означает, что данная оценка имеет наименьший разброс значений.

Теорема 6. Пусть случайная величина X обладает конечной дисперсией: $DX = \sigma^2 < \infty$. Оценка $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ является несмещенной и состоятельной оценкой параметра $a = MX$.

Доказательство. Так как случайная величина X_k имеет ту же функцию распределения, что и случайная величина X, то $MX_k = MX = a$, $DX_k = DX = \sigma^2$.

В силу линейности математического ожидания $M\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} MX_k = \frac{1}{n}\cdot na = a$.

Несмещенность оценки доказана.

Докажем теперь состоятельность данной оценки. Так как случайные величины $X_1, .., X_n$ независимы, то по свойству дисперсии имеем:

$$D\overline{X}_n = D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} DX_k = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Используя второе неравенство Чебышева, получим:

$$P\left(\left|X_n - a\right| \geq \varepsilon\right) = P\left(\left|\overline{X}_n - M\overline{X}_n\right| \geq \varepsilon\right) \leq \frac{D\overline{X}_n}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon_n^2} \xrightarrow{n\to\infty} 0.$$

Очевидно, что $\left\{\left|X_n - a\right| \geq \varepsilon\right\} = \left\{\overline{\left|X_n - a\right| < \varepsilon}\right\}$.

$$0 = \lim_{n\to\infty} P\left[\left\{\overline{\left|\overline{X}_n - a\right| < \varepsilon}\right\}\right] = \lim_{n\to\infty}\left[1 - P\left(\left|\overline{X}_n - a\right| < \varepsilon\right)\right] = 1 - \lim_{n\to\infty} P\left(\left|\overline{X}_n - a\right| < \varepsilon\right) \Rightarrow$$

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - a\right| < \varepsilon\right) = 1.$$

Последнее равенство и доказывает состоятельность оценки.

□

<u>*Теорема 7.*</u> Справедливы следующие утверждения:

1) Оценка $\hat{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$ параметра $\sigma^2 = DX$ является несмещенной оценкой.

2) Если существует математическое ожидание от $MX^4 < \infty$, то данная оценка состоятельна.

<u>*Доказательство.*</u> 1) Сформулируем и докажем вспомогательную лемму.

<u>*Лемма 1.*</u> Пусть $Y_k = X_k - C$, где C=const, $k = \overline{1,n}$. Тогда $\hat{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_k - \overline{Y}\right)^2$.

<u>*Доказательство.*</u> В выражении $\hat{S}_n^2$ подставим $Y_k + C$ вместо X_k. Имеем:

$$X_k - \overline{X} = Y_k + C - \frac{1}{n}\sum_{k=1}^{n}\left(Y_k + C\right) = Y_k + C - \frac{1}{n}\sum_{k=1}^{n}Y_k - \frac{1}{n}\sum_{k=1}^{n}C = Y_k - \overline{Y}.$$

□

Продолжим доказательство первой части теоремы. Положим $C = a = MX_n$. Тогда $MY_k = M\left(X_k - a\right) = MX_k - a = a - a = 0$, $MY_k^2 = DY_k = DX_k = \sigma^2$. Преобразуем

$$\hat{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_k - \overline{Y}\right)^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_k^2 - 2Y_k\overline{Y} + \overline{Y}\right) = \frac{1}{n-1}\sum_{k=1}^{n}Y_k^2 - \frac{2\overline{Y}}{n-1}\sum_{k=1}^{n}Y_k + \frac{1}{n-1}n\overline{Y}^2 =$$

$$= \frac{1}{n-1}\sum_{k=1}^{n}Y_k^2 - \frac{n}{n-1}\overline{Y}^2.$$

$$M\hat{S}_n^2=\frac{1}{n-1}\sum_{k=1}^{n}MY_k^2-\frac{n}{n-1}M\overline{Y}^2=\frac{n\sigma^2}{n-1}-\frac{n}{n-1}M\overline{Y}^2.$$ **В**ычислим $M\overline{Y}^2$:

$$M\overline{Y}^2=M\left(\frac{1}{n}\sum_{k=1}^{n}Y_k\right)^2=\frac{1}{n}M\left(\sum_{k=1}^{n}Y_k\right)^2=\frac{1}{n^2}M\left(\sum_{i,j=1}^{n}Y_iY_j\right)=\frac{1}{n^2}M\left(\sum_{k=1}^{n}Y_k^2+\sum_{i,j=1}^{n}Y_iY_j\right)=$$

$$=\frac{1}{n^2}M\left(\sum_{k=1}^{n}Y_k^2\right)+\frac{1}{n^2}M\left(\sum_{\substack{i,j=1\\ i\neq j}}^{n}Y_iY_j\right)=\frac{1}{n^2}n\sigma^2+\frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}M\left(Y_iY_j\right)=\frac{\sigma^2}{n}\oplus Y_1,Y_2,\ldots,Y_n-\textit{независимы}$$

$$\oplus\sum_{\substack{i,j=1\\ i\neq j}}^{n}MY_iMY_j=\frac{\sigma^2}{n}.$$

$$M\hat{S}_n^2=\frac{n\sigma^2}{n-1}-\frac{\sigma^2}{n-1}=\frac{\sigma^2}{n-1}(n-1)=\sigma^2.$$

Итак, несмещённость данной оценки доказана.

Лекция 5 по математической статистике

2) Для доказательства состоятельности сформулируем следующую лемму.

Лемма 2. Если $MX^4<\infty$, то существует такое число $c>0$, что $D\hat{S}_n^2\leq\frac{c}{n}$.

Применим второе неравенство Чебышева:

$$P\left(\left|\hat{S}_n^2-\sigma^2\right|\geq\varepsilon\right)=P\left(\left|\hat{S}_n^2-MS_n^2\right|\geq\varepsilon\right)\leq\frac{D\hat{S}_n^2}{\varepsilon^2}=\frac{c}{\varepsilon^2 n}\to 0,\quad \text{при } n\to\infty.$$

Действуя далее аналогично, как и при доказательстве теоремы 6, получаем состоятельность данной оценки дисперсии.

□

Пример 9. Найти несмещенную оценку дисперсии случайной величины X на основании данного распределения выборки:

x_i	2	7	9	10
n_i	8	14	10	18

Решение.

Находим выборочную среднюю $\overline{X}=\frac{2\cdot 8+7\cdot 14+9\cdot 10+10\cdot 18}{50}=7{,}68$.

Для вычисления выборочной дисперсии воспользуемся формулой: $\hat{D}X=\overline{X}^2-\left(\overline{X}\right)^2$.

$\overline{X}^2=\frac{4\cdot 8+49\cdot 14+81\cdot 10+100\cdot 18}{50}=66{,}56$, $\hat{D}X=66{,}56-(7{,}68)^2=7{,}58$. Находим несмещенную оценку дисперсии («исправленную» выборочную дисперсию): $\hat{S}^2=\frac{n}{n-1}\hat{D}X=\frac{50}{49}\cdot 7{,}68=7{,}73$.

□

Пример 10. Монету подбрасывают n раз. Вероятность выпадения герба при каждом подбрасывании равна p. В ходе опыта монета выпала гербом n_1 раз.

Показать несмещенность оценки $\hat{\theta} = \dfrac{n_1}{n}$ вероятности $\theta = p$ выпадения герба в каждом опыте.

Решение. Число успехов n_1 имеет биномиальное распределение.
Тогда $M(n_1) = n \cdot p$, $D(n_1) = npq = np(1-p)$.
Следовательно, $M\hat{\theta} = M\left(\dfrac{n_1}{n}\right) = \dfrac{1}{n} \cdot M(n_1) = \dfrac{1}{n} \cdot n \cdot p = p = \hat{\theta}$, что доказывает несмещенность оценки $\hat{\theta} = \dfrac{n_1}{n}$.

□

Упражнение. Исследовать на несмещённость и состоятельность следующую оценки дисперсии: $\hat{S}_a^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}(x_k - a)^2$, a - теоретическое значение математического ожидания.

Лекция 6

Методы нахождения точечных оценок.

Метод наибольшего правдоподобия

Пусть $F(\theta, x) = P(X_k \le x)$ — функция распределения случайной величины, имеющей плотность $p(\theta, x)$.

Тогда функция $L = L(X_1, X_2, ..., X_n, \theta) = P_{X_1}(\theta) \cdot P_{X_2}(\theta) \cdot ... \cdot P_{X_n}(\theta)$ является функцией только параметра θ. По методу наибольшего правдоподобия за оценку параметра θ принимают значение аргумента $\theta = \theta(X_1, X_2, ..., X_n)$, при котором функция L имеет максимальное значение, т.е. выбирают такое значение θ, при котором вероятность получения выборки именно этих значений наибольшая.

Рассмотрим уравнение $\dfrac{d \ln L}{d\theta} = 0$.

Решением этого уравнения будем называть только корни, зависящие от выборки $\theta > \theta(X_1, X_2, \ldots, X_n)$, а само решение будем называть **оценкой наибольшего правдоподобия.**

Если нужно получить несколько оценок $\theta_1, \theta_2, ..., \theta_S$, то нужно решить следующую систему уравнений:

$$\begin{cases} \dfrac{d \ln L}{d\theta_1} = 0 \\ \ldots\ldots\ldots\ldots \\ \dfrac{d \ln L}{d\theta_S} = 0 \end{cases}$$

<u>*Пример 11.*</u>

Рассмотрим случайные величины, распределенные по нормальному закону:

$X_k, k = \overline{1, n} \quad a = MX_k \quad b = DX_k$

Получим их оценку:

$$L = L(X_1, X_2, ..., X_n, a, b) = \left(\frac{1}{\sqrt{2\pi b}} \right)^n \cdot e^{-\sum_{k=1}^{n} \left(\frac{X_k - a}{2b} \right)^2}$$

$$\ln L = -\frac{n}{2}(\ln 2\pi + \ln b) - \frac{1}{2b} \sum_{k=1}^{n} (X_k - a)^2$$

$$\frac{d \ln L}{da} = \frac{1}{b} \sum_{k=1}^{n} (X_k - a) \qquad \frac{d \ln L}{db} = \frac{-n}{2b} + \frac{1}{2b^2} \sum_{k=1}^{n} (X_k - a)^2$$

$$\sum_{k=1}^{n} (X_k - a) = 0 \ , \quad -nb + \sum_{k=1}^{n} (X_k - \overline{X})^2 = 0,$$

$$\sum_{k=1}^{n} X_k = \sum_{k=1}^{n} a \quad , \qquad \hat{b} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 , \qquad \hat{a} = \overline{X} .$$

<u>*Пример 12.*</u> Найти оценку наибольшего правдоподобия для вероятности успеха в схеме Бернулли $X_1,\ldots,X_n : X_k = 0$ *или* 1.

Составим функцию правдоподобия

$$L = L\left(X_1, X_2, \ldots, X_{n,p}\right) = \begin{cases} \prod_{k=1}^{n} p_n^{X_k} , & \text{если } X_k = 1, \\ \prod_{k=1}^{n} p_n^{X_k} \cdot (1 - p)^{1-X_k} , & \text{если } X_k = 0 . \end{cases}$$

$$\ln L = \sum_{k=1}^{n}\left(X_k \cdot \ln p + \left(1 - X_k\right) \cdot \ln(1 - p)\right).$$

$$\frac{d \ln L}{dp} = \sum_{k=1}^{n}\left(\frac{X_k}{p} - \frac{1 - X_k}{1 - p}\right) = \frac{1}{p}\sum_{k=1}^{n} X_k - \frac{n}{1 - p} + \frac{1}{1 - p}\sum_{k=1}^{n} X_k =$$

$$= \frac{1}{p(1 - p)}\left((1 - p) \cdot \sum_{k=1}^{n} X_k - np + p\sum_{k=1}^{n} X_k\right) = 0.$$

$$\sum_{k=1}^{n} X_k - np = 0 \Rightarrow \hat{p} = \hat{x} \qquad \text{или} \qquad \mu_n = \sum_{k=1}^{n} X_k , \hat{p} = \frac{\mu_n}{n} .$$

Метод наименьших квадратов

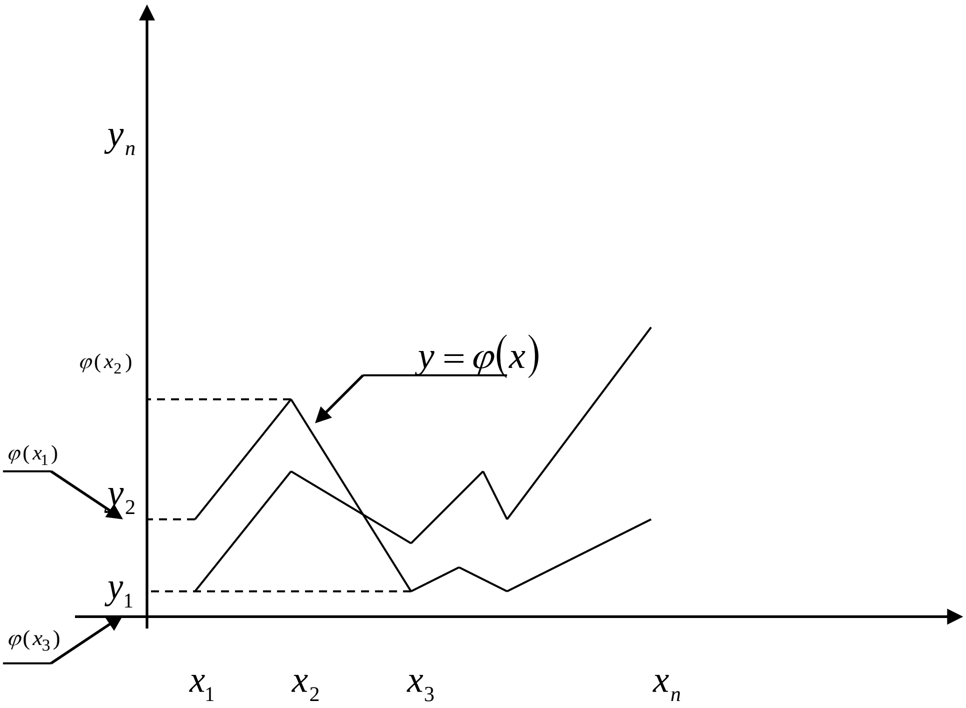
Рассмотрим следующую задачу. Пусть существует следующая функциональная зависимость y=$\varphi(X)$, которая нам неизвестна. Допустим, что нам точно известны некоторые аргументы $X_1, X_2, \ldots, X_n$ и приближенные значения заданной функции данных аргументов: $y_1, y_2, \ldots, y_n$. Считается, что эти приближенные значения функции получены в результате моделирования значений функции с использованием датчика случайных чисел. Используем естественную связь приближенных значений и точных значений:

$$y_k = \varphi(X_k) + \sigma_k ,$$

где σ_k - ошибка измерений.

Естественно считать, что $\sigma_1, \sigma_2, \ldots, \sigma_n$ – совокупность независимых случайных величин, удовлетворяющих условию: $M(\sigma_k) = 0$, при $\forall\ k = 1,2,\ldots,n.$

Будем предполагать, что $D(\sigma_k) = \sigma^2$, $k = 1,2,\ldots,n.$ Если нет дополнительных сведений от функции, то нужно построить следующий график:



Однако полученная кривая плохо приближает график неизвестной функции y=$\varphi(X)$. Поэтому чаще всего при решении задачи отыскания графика функции y=$\varphi(X)$ считаются известными некоторые свойства функции (например: свойство гладкости, или то, что функция является многочленом). В этом случае для определения функции y=$\varphi(X)$ и применяется ***метод наименьших квадратов***, заключающийся в следующем. Нарисуем на данном графике истинную функцию:

$$Q=(y_1-\varphi(X_1))^2+(y_2-\varphi(X_2))^2+\ldots+(y_n-\varphi(X_n))^2+\ldots+(y_n-\varphi(X_n))^2=$$

$$=\sum_{k=1}^{n}(y_k-\varphi(X_k))^2.$$

Степень близости приближенных значений и точных значений функции будет наибольшей, когда функция Q принимает наименьшие значения. Таким образом, мы должны решить задачу нахождения наименьшего значения функции Q. В результате решения этой задачи мы получим оценки параметров, характеризующих функцию y=$\varphi(X)$. При некоторых условиях на функцию φ эти оценки будут состоятельны.

<u>Случай 1.</u> Пусть известно, что функция y=$\varphi(X)=AX+B$ с неизвестными A и B. Пользуясь методом наименьших квадратов, найдём оценки $\hat{A}$ и $\hat{B}$ этих коэффициентов. Составим для этого функцию Q.

$Q(A,B)=\sum_{k=1}^{n}(y_k-AX_k-B)^2$. Надо вычислить значения с таким расчетом, чтобы функция $Q(A,B)$ принимала наименьшие значения. Вычислим частные производные и приравняем их к нулю.

$$\frac{dQ}{dA}=\sum_{k=1}^{n}2\left(y_k-AX_k-B\right)\cdot\left(-X_k\right)=0.$$

$$\frac{dQ}{dB}=\sum_{k=1}^{n}2\left(y_k-AX_k-B\right)\cdot\left(-1\right)=0.$$

$$\begin{cases}\sum\limits_{k=1}^{n}X_k y_k-A\sum\limits_{k=1}^{n}X_k^2-B\sum\limits_{k=1}^{n}X_k=0\\ \sum\limits_{k=1}^{n}y_k-A\sum\limits_{k=1}^{n}X_k-Bn=0.\end{cases}$$

<u>*Лемма.*</u> Пусть ось OY можно перевернуть параллельно себе таким образом, что в новой системе координат сумма значений всех данных абсцисс будет равна нулю.

□

Вернемся к рассуждениям. Применяя лемму, можно считать, что в нашей системе $\sum\limits_{k=1}^{n}X_k=0$. При этом система принимает вид:

$$\begin{cases}\sum\limits_{k=1}^{n}X_k y_k-A\sum\limits_{k=1}^{n}X_k^2=0\\ \sum\limits_{k=1}^{n}y_k-Bn=0\end{cases}$$

Следовательно, $\hat{A}=\dfrac{\sum\limits_{k=1}^{n}X_k y_k}{\sum\limits_{k=1}^{n}X_k^2}$, $\hat{B}=\dfrac{1}{n}\sum\limits_{k=1}^{n}y_k$.Таким образом, мы получили оценку параметров A и B.

Решим вопрос о том, являются ли эти оценки несмещенными и состоятельными. Для этого выполним в данной системе замену $y_k=Ax+B+\delta_k$.

$$\hat{A}=A+\frac{\sum\limits_{k=1}^{n}X_k\delta_k}{\sum\limits_{k=1}^{n}X_k^2}\qquad\qquad \hat{B}=B+\frac{1}{n}\sum_{k=1}^{n}\delta_k$$

Вычислим математические ожидания и дисперсии случайных величин, входящих в состав $\hat{A}$ и $\hat{B}$:

$$M\left(\frac{\sum\limits_{k=1}^{n}X_k\delta_k}{\sum\limits_{k=1}^{n}X_k^2}\right)=\frac{1}{\sum\limits_{k=1}^{n}X_k^2}\cdot\sum_{k=1}^{n}X_k\cdot M\delta_k=0.$$ Аналогично, $M\left(\dfrac{1}{n}\sum\limits_{k=1}^{n}\delta_k\right)=0$.

$D\left(\dfrac{\sum_{k=1}^{n} X_k \delta_k}{\sum_{k=1}^{n} X_k^2}\right) = \dfrac{1}{\left(\sum_{k=1}^{n} X_k^2\right)^2} \cdot \sum_{k=1}^{n} X_k^2$. Так как члены случайных величин попарно независимы, то $D\left(\dfrac{\sum_{k=1}^{n} X_k \delta_k}{\sum_{k=1}^{n} X_k^2}\right) = \dfrac{1}{\left(\sum_{k=1}^{n} X_k^2\right)^2} \cdot \sum_{k=1}^{n} X_k^2 \cdot \delta^2 = \dfrac{\delta^2}{\sum_{k=1}^{n} X_k^2}$,

$$D\left(\frac{1}{n}\sum_{k=1}^{n}\delta_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\underbrace{D\delta_k}_{\delta^2} = \frac{1}{n^2}\cdot\delta^2\cdot n = \frac{\delta^2}{n} \Rightarrow \quad M\hat{A} = A; \quad M\hat{B} = B.$$

$D\hat{A} = DA + D\left(\dfrac{\sum_{k=1}^{n} X_k \delta_n}{\sum_{k=1}^{n} X_k^2}\right) = \dfrac{\delta^2}{\sum_{k=1}^{n} X_k^2}$. Аналогично $D\hat{B} = \dfrac{\delta^2}{n}$.

Из приведенных выкладок вытекает справедливость следующей теоремы:

<u>*Теорема 8*</u>**.** Если неизвестная функция $\varphi(x)$ имеет вид $\varphi(x) = Ax + B$ и параметры A и B оценены с помощью метода наименьших квадратов величинами $\hat{A}$ и $\hat{B}$, то эти оценки являются несмещенными, а оценка $\hat{B}$ является состоятельной. Оценка $\hat{A}$ состоятельна, если ряд расходится.

Оценим теперь параметры A и B, используя принцип наибольшего правдоподобия. При этом будем предполагать, что случайные величины δ_k имеют нормальное распределение.

$y_k = Ax_k + B + \delta_k$, k=1, 2,...,n

Ясно, что случайные величины y_k также имеют нормальное распределение. Имеем:

$My_k = Ax_k + B$, $Dy_k = \Delta\delta_k - \delta^2$. Тогда $P_{y_k} = \dfrac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{(y_k - Ax_k - B)^2}{2\sigma^2}}$.

Составим функцию правдоподобия

$$L = \prod_{k=1}^{n} Py_k = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - Ax_k - B)^2}$$

$$\ln L = \ln\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - Ax_k - B)^2$$

$$\begin{cases} \dfrac{\delta \ln L}{\delta A} = \dfrac{1}{2\sigma}\cdot 2\sum_{k=1}^{n}(y_x - Ax_k - B)\cdot(-x_k) = 0 \\ \dfrac{\delta \ln L}{\delta B} = \dfrac{1}{2\sigma}\cdot 2\sum_{k=1}^{n}(y_x - Ax_k - B)\cdot(-1) = 0 \end{cases}$$

Получается та же система, что и при выведении метода наименьших квадратов.

Случай 2. Пусть $y=\varphi(x)=Ax^2+Bx+C$. Нужно оценить параметры *A, B, C*.

$X_k=\dfrac{2k}{n}$, *k=1,...,n.*

Смоделируем измерения чисел y_k по формулам: $y_k=-2X_k^2+10X_k-8+\delta_k, M\delta_k=0$, $D\delta_k=0{,}04$.

При этом числа *x* взяты из датчика случайных чисел.

Применим метод наименьших квадратов для оценки параметров *A, B, C*. Составим

$$Q(A,B,C)=\sum_{k=1}^{n}\left(y_x-AX_k^2-BX_k-C\right)^2$$

$$\begin{cases}\dfrac{\delta Q}{\delta A}=-2\sum\limits_{k=1}^{n}X_k^2\left(y_x-AX_k^2-BX_k-C\right)=0\\ \dfrac{\delta Q}{\delta B}=-2\sum\limits_{k=1}^{n}X_k\left(y_x-AX_k^2-BX_k-C\right)=0\\ \dfrac{\delta Q}{\delta C}=-2\sum\limits_{k=1}^{n}\left(y_x-AX_k^2-BX_k-C\right)=0\end{cases}$$

Решив эту систему с 3-мя неизвестными, получим:

$n=25$	$n=100$
$\hat{A}=-1.871$	$\hat{A}=-2.046$
$\hat{B}=9.510$	$\hat{B}=10.200$
$\hat{C}=-7.627$	$\hat{C}=-8.218$

Лекция 7

Интервальные оценки параметров распределения

Точечные оценки неизвестного параметра θ хороши в качестве первоначальных результатов обработки наблюдений. Их недостаток состоит в том, что неизвестна точность оценивания параметра. Поэтому и возникает задача о приближении параметра θ не одним числом, а целым интервалом.

Если известны закон распределения случайной величины, являющийся оценкой, и ее дисперсия, то можно указать границы, в пределах которых находится неизвестное значение параметра с большей вероятностью. Если эти границы зависят от значения неизвестного параметра, то ими пользоваться нельзя.

Рассмотрим выборку $X_1, X_2, ..., X_n : P(X_i \le x) = F(\theta, x)$, где θ — неизвестный параметр. Пусть удалось найти значение параметра θ $\bar{\theta}(X_1, X_2, ..., X_n)$, удовлетворяющее условию $\theta < \bar{\theta}$ для любых значений $X_1, X_2, ..., X_n$. Предположим, что для любых значений θ выполняется соотношение:

$$P\left(\underline{\theta}(X_1, X_2, ..., X_n) < \theta < \bar{\theta}(X_1, X_2, ..., X_n)\right) = 1 - 2\alpha,$$

где α — некоторое число.

Таким образом, значение параметра не зависит от значения случайной величины X и не связано со значением параметра θ.

<u>*Определение 37.*</u> Интервал $(\underline{\theta}; \bar{\theta})$ называется ***доверительным интервалом*** θ с доверительной вероятностью (надежностью) $\gamma = 1 - 2\alpha$.

Доверительные интервалы для параметров нормального распределения

Построим доверительные интервалы для параметров нормального распределения, то есть когда выборка производится из генеральной совокупности, имеющей нормальное распределение с параметрами a и σ.

1) Доверительный интервал для математического ожидания при известной дисперсии

Пусть выборка $X_1, X_2, ..., X_n$ распределена по нормальному закону с известным среднеквадратичным отклонением σ и требуется найти доверительный интервал для математического ожидания a с надежностью γ.

Выборочное среднее $\bar{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ — тоже распределено по нормальному закону с параметрами $M\bar{X} = a$, $D\bar{X} = \dfrac{\sigma^2}{n}$. Найдем такое число ε, чтобы выполнялось равенство

$$P\left(\bar{X} - \varepsilon < a < \bar{X} + \varepsilon\right) = \gamma.$$

Введем такую величину: $\dfrac{\overline{X}-a}{\dfrac{\sigma}{\sqrt{n}}}$.

Математическое ожидание этой случайной величины равно 0, среднеквадратичное отклонение равно 1, и ее распределение не зависит от a. Найдем для заданной вероятности γ такое число U_γ, что выполняется равенство:

$$P\left\{\left|\frac{\overline{X}-a}{\dfrac{\sigma}{\sqrt{n}}}\right|<U_\gamma\right\}=\gamma,$$

где U_γ является решением уравнения $2\Phi(U_\gamma)=\gamma$.

Выполним следующие тождественные преобразования:

$$\left|\frac{\overline{X}-a}{\dfrac{\sigma}{\sqrt{n}}}\right|<U_\gamma \Rightarrow \left|\overline{X}-a\right|<U_\gamma\cdot\frac{\sigma}{\sqrt{n}} \Rightarrow -U_\gamma\cdot\frac{\sigma}{\sqrt{n}}<\overline{X}-a<U_\gamma\cdot\frac{\sigma}{\sqrt{n}} \Rightarrow$$

$$\Rightarrow -\overline{X}-U_\gamma\cdot\frac{\sigma}{\sqrt{n}}<-a<-\overline{X}+U_\gamma\cdot\frac{\sigma}{\sqrt{n}} \Rightarrow \overline{X}-U_\gamma\cdot\frac{\sigma}{\sqrt{n}}<a<\overline{X}+U_\gamma\cdot\frac{\sigma}{\sqrt{n}}.$$

Таким образом, вероятность

$$P\left(\overline{X}-U_\gamma\cdot\frac{\sigma}{\sqrt{n}}<a<\overline{X}+U_\gamma\cdot\frac{\sigma}{\sqrt{n}}\right)=\gamma,$$

поэтому

$$\varepsilon=U_\gamma\cdot\frac{\sigma}{\sqrt{n}},$$

а интервальная оценка для математического ожидания имеет вид:

$$\left(\overline{X}-U_\gamma\cdot\frac{\sigma}{\sqrt{n}};\overline{X}+U_\gamma\cdot\frac{\sigma}{\sqrt{n}}\right) \tag{24}$$

2) Доверительный интервал для математического ожидания при неизвестной дисперсии

Решим задачу построения интервальной оценки для математического ожидания нормального распределения при условии неизвестной дисперсии. Пусть случайная величина $X=N(a,\sigma)$, причем неизвестными являются и a, и σ. Вычислим по $X_1,X_2,\ldots,X_n$ выборочное среднее $\overline{X}$ и оценку $S^2=\dfrac{1}{n-1}\sum\limits_{k=1}^{n}\left(X_k-\overline{X}\right)^2$ дисперсии σ. Используя заданную доверительную вероятность γ, найдем такое число ε, чтобы выполнялось равенство

$$P\left(\overline{X}-\varepsilon<a<\overline{X}+\varepsilon\right)=\gamma \tag{25}$$

Рассмотрим величину

$$t_{n-1} = \frac{\overline{X} - a}{S} \cdot \sqrt{n} \qquad (26)$$

которая не зависит от параметров a и σ и распределена по закону Стьюдента с (n-1) степенями свободы.

Определим величину $t_{\gamma,n-1}$ как решение уравнения:

$$P\left(\left|t_{n-1}\right| < t_{\gamma,n-1}\right) = \gamma \qquad (27)$$

Тогда получим

$$P\left(\left|\frac{\overline{X} - a}{S/\sqrt{n}}\right| < t_{\gamma,n-1}\right) = \gamma \qquad (28)$$

Выполним следующие тождественные преобразования:

$$\left(\left|\frac{\overline{X} - a}{S/\sqrt{n}}\right| < t_{\gamma,n-1}\right) \Rightarrow \left|\overline{X} - a\right| < \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} \Rightarrow -\frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} < \overline{X} - a < \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} \Rightarrow$$

$$\Rightarrow -\overline{X} - \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} < -a < -\overline{X} + \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} \Rightarrow \overline{X} - \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} < a < \overline{X} + \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}}.$$

Итак,

$$P\left(\overline{X} - \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} < a < \overline{X} + \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}}\right) = \gamma \qquad (29)$$

Сравнивая последнее равенство с равенством (25), получаем, что

$$\varepsilon = \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}} \qquad (30)$$

Следовательно, с вероятностью γ можно заключить, что выборочное среднее $\overline{X}$ дает значение неизвестного математического ожидания с точностью ε, а доверительный интервал определяется как

$$\left(\overline{X} - \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}}, \overline{X} + \frac{S \cdot t_{\gamma,n-1}}{\sqrt{n}}\right) \qquad (31)$$

3) Доверительный интервал для неизвестной дисперсии и среднего квадратического отклонения

Пусть случайная величина $X = N(a,\sigma)$, причем неизвестным является σ. Вычислим по $X_1, X_2, \ldots, X_n$ выборочное среднее $\overline{X}$ и оценку $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$ дисперсии σ. В качестве оценки неизвестного среднего квадратического отклонения возьмем $S = \sqrt{S^2}$. Используя заданную доверительную вероятность γ, найдем такое число ε, чтобы выполнялось равенство

$$P(S - \varepsilon < \sigma < S + \varepsilon) = \gamma \qquad (32)$$

Рассмотрим величину

$$\chi_{n-1}^{2}=\frac{(n-1)\cdot S^{2}}{\sigma^{2}} \tag{33}$$

которая имеет χ^2- распределение с $(n-1)$-й степенью свободы. Найдем для заданных γ и $(n-1)$ такое число q_γ, при котором верно равенство

$$P\left(\frac{n-1}{(1+q_\gamma)^2}<\chi_{n-1}^{2}<\frac{n-1}{(max(0;1-q_\gamma))^2}\right)=\gamma \tag{34}$$

Учитывая (33), получим

$$P\left(\frac{n-1}{(1+q_\gamma)^2}<\frac{(n-1)\cdot S^{2}}{\sigma^{2}}<\frac{n-1}{(max(0;1-q_\gamma))^2}\right)=\gamma \tag{35}$$

Выполним следующие тождественные преобразования:

$$\frac{1}{(1+q_\gamma)^2}<\frac{S^2}{\sigma^2}<\frac{1}{(max(0;1-q_\gamma))^2}\Rightarrow\frac{1}{S^2(1+q_\gamma)^2}<\frac{1}{\sigma^2}<\frac{1}{S^2\cdot(max(0;1-q_\gamma))^2}\Rightarrow$$

$$\Rightarrow S^2(1+q_\gamma)^2>\sigma^2>S^2\cdot(max(0;1-q_\gamma))^2\Rightarrow S\cdot(max(0;1-q_\gamma))<\sigma<S\cdot(1+q_\gamma).$$

Таким образом, вероятность

$$P\left(S\cdot(max(0;1-q_\gamma))<\sigma<S\cdot(1+q_\gamma)\right)=\gamma \tag{36}$$

или

$$P\left(max(0;S\cdot(1-q_\gamma))<\sigma<S\cdot(1+q_\gamma)\right)=\gamma \tag{36'}$$

Следовательно,

$$\varepsilon=S\cdot q_\gamma \tag{37}$$

Итак, с вероятностью γ можно утверждать, что интервал

$$\left(max(0;\,S-S\cdot q_\gamma);S+S\cdot q_\gamma\right) \tag{38}$$

накроет неизвестное среднее квадратическое отклонение σ.

С такой же вероятностью можно утверждать, что интервал

$$\left((max(0;\,S-S\cdot q_\gamma))^2;(S+S\cdot q_\gamma)^2\right) \tag{39}$$

накроет неизвестную дисперсию σ^2.

Лекция 8

Проверка статистических гипотез

Одним из часто встречающихся применений статистических методов на практике является решение вопроса о том, должно ли быть принято или отвергнуто предположение (гипотеза) относительно генеральной совокупности (случайной величины) на основании данной выборки. Например, новое правило поступления в вуз испытано на определенном числе абитуриентов. Логично выяснить, можно ли сделать обоснованный вывод о его эффективности по сравнению с предыдущим правилом поступления по данным результатам?

Определение 38. ***Статистической гипотезой*** называется любое высказывание (предположение) о генеральной совокупности, проверяемое по выборке (по результатам наблюдений).

Определение 39. ***Проверкой гипотезы*** называется процедура сопоставления высказанного предположения (гипотезы) с выборочными данными .

Статистические гипотезы делятся на параметрические (гипотезы о параметрах распределения известного вида) и непараметрические (о виде неизвестного распределения).

Основываясь на выборочных данных и учитывая условия конкретной задачи, выдвигают гипотезу H_0 в качестве основной (***нулевой***), а другую гипотезу H_1, являющуюся ее логическим отрицанием, – в качестве конкурирующей (***альтернативной***). Например, если нулевая гипотеза состоит в том, что математическое ожидание равно 10 ($H_0: MX = 10$), то в качестве альтернативной можно рассматривать одну из следующих гипотез: $MX < 10,\ MX > 10,\ MX = 5$.

Имея две гипотезы H_0и H_1, надо на основе выборки $X_1, X_2, ..., X_n$ принять либо нулевую гипотезу H_0, либо альтернативную гипотезу H_1. Правило, согласно которому принимается или отклоняется гипотеза H_0, называется ***критерием*** проверки гипотезы H_0.

При проверке гипотезы может быть принято неправильное решение, то есть, допущены ошибки 1-го и 2-го рода. ***Ошибка 1-го рода*** состоит в том, что отвергается нулевая гипотеза H_0, когда на самом деле она верна. Вероятность ошибки 1-го рода называется ***уровнем значимости критерия*** и обозначается через α. ***Ошибка 2-го рода*** состоит в том, что принимается нулевая гипотеза H_0, когда на самом деле верна гипотеза H_1. Вероятность ошибки 2-го рода обозначается через β. Величина $1-\beta$ называется ***мощностью критерия***. Обобщим сказанное в виде таблицы.

Гипотеза H_0	Отвергается	Принимается
верна	Ошибка 1-го рода, ее вероятность равна α	Правильное решение, его вероятность равна $1-\alpha$
неверна	Правильное решение, его вероятность равна $1-\beta$	Ошибка 2-го рода, ее вероятность равна β

Схема проверки гипотез сводится к следующему:

Этап 1. Располагая выборкой $X_1, X_2, ..., X_n$, формулируют нулевую гипотезу H_0 и альтернативную H_1.

Этап 2. В каждом конкретном случае подбирают ***статистику критерия*** (функцию выборки, формирующуюся на основании результатов выборки $X_1, X_2, ..., X_n$) $T_n = T(X_1, X_2, ..., X_n)$, обычно из перечисленных ниже: U – нормальное распределение, χ^2- распределение Пирсона, t – распределение Стьюдента, F – распределение Фишера – Снедекора.

Этап 3. По статистике критерия T_n и заданному уровню значимости α определяют критическую область S (и $\overline{S}$). Для ее отыскания достаточно найти критическую точку $t_{кр}$, то есть границу, отделяющую область S от $\overline{S}$. Границы областей определяются из соотношений:

- для правосторонней критической области: $P(T_n > t_{кр}) = \alpha$.
- для левосторонней критической области: $P(T_n < t_{кр}) = \alpha$.
- для двусторонней критической области: $P(T_n < t_{кр}^{л}) = P(T_n > t_{кр}^{п}) = \dfrac{\alpha}{2}$.

Для каждого критерия критическую точку, удовлетворяющую приведенным соотношениям, находят по соответствующим таблицам.

Этап 4. Для полученной реализации выборки $x_1, x_2, ..., x_n$ вычисляют значение критерия $T_{набл} = T(x_1, x_2, ..., x_n) = t$.

Этап 5. Если $t \in S$, то нулевую гипотезу H_0 отвергают и принимают альтернативную H_1; если же $t \in \overline{S}$, то нулевую гипотезу H_0 принимают.

Проверка гипотезы о виде закона распределения. Критерии согласия

Во многих практических задачах модель закона распределения заранее неизвестна, поэтом возникает вопрос выбора модели, согласующейся с результатами наблюдения над случайной величиной X.

Предположим, что неизвестная функция распределения $F_X(x)$ случайной величины X имеет определенную модель $F_{теор}(x)$, то есть сформулируем гипотезу $H_0 : F_X(x) = F_{теор}(x)$. Тогда в качестве альтернативной выдвинем гипотезу $H_1 : F_X(x) \neq F_{теор}(x)$. Требуется сделать вывод: согласуются ли данные наблюдений с высказанным предположением?

<u>*Определение 40.*</u> ***Критерием согласия*** называется критерий, с помощью которого проверяется гипотеза о предполагаемом виде закона распределения (о согласовании предполагаемого вида распределения с опытными данными на основании выборки).

Критерий χ^2 Пирсона

Для проверки гипотезы H_0 поступим следующим образом. Разобъем всю область значений случайной величины X на m интервалов $\Delta_1, \Delta_2, ..., \Delta_m$ и подсчитаем вероятности p_i $(i = 1,2,...,m)$ попадания случайной величины X в интервал Δ_i по формуле

$$P(\alpha \le X \le \beta) = F_{теор}(\beta) - F_{теор}(\alpha) \tag{40}$$

Тогда теоретическое число значений случайной величины X, попавших в интервал Δ_i, можно вычислить по формуле

$$n_i^{теор} = n \cdot p_i \tag{41}$$

Таким образом, получим вариационный ряд распределения и теоретический ряд распределения. Если эмпирические частоты сильно отличаются от теоретических, то проверяемую гипотезу H_0 отвергаем, в противном случае – принимаем ее.

В качестве критерия, характеризующего степень расхождения между эмпирическими и теоретическими частотами, К.Пирсон предложил статистику

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(n_i - n_i^{теор}\right)^2}{n_i^{теор}} \tag{42}$$

При $n \to \infty$ эта величина имеет χ^2 – распределение с $k = m - r - 1$ степенями свободы, где m – число интервалов выборки, r – число параметров предполагаемого распределения. Например, в случае нормального распределения оценивают два параметра a и σ, поэтому $r = 2$.

Схема применения критерия χ^2 Пирсона сводится к следующему:

1. По формуле (42) вычисляем $\chi^2_{набл}$ – выборочное значение статистики критерия.
2. Задав уровень значимости α критерия, по таблице χ^2 – распределения находим критическую точку (квантиль) $\chi^2_{крит}$.
3. Если $\chi^2_{набл} \le \chi^2_{крит}$, то гипотеза H_0 не противоречит данным наблюдений; в противном случае если $\chi^2_{набл} > \chi^2_{крит}$, то гипотезу H_0 отвергаем.

Необходимым условием применения критерия χ^2 Пирсона является выполнение соотношения $n_i^{теор} \ge 5$. Если для какой-то группы выборки оно не выполняется, такую группу объединяют с соседней и соответственно уменьшают число групп.

Критерий Колмогорова

Критерий Колмогорова является наиболее простым критерием проверки гипотезы о модели закона распределения. Он связывает эмпирическую функцию распределения $\hat{F}_n(x)$ с функцией распределения $F(x)$ непрерывной случайной величины X.

Рассмотрим $x_1, x_2, ..., x_n$ – конкретную выборку из распределения с неизвестной функцией распределения $F(x)$. Пусть $\hat{F}_n(x)$ – эмпирическая функция распределения. Сформулируем простую гипотезу H_0: $F(x) = F_0(x)$, в качестве альтернативной выдвинем гипотезу H_1: $F(x) \neq F_0(x)$.

Согласно критерию Колмогорова вводится в рассмотрение функция

$$D_n = \max_{-\infty < x < \infty} \left| \hat{F}_n(x) - F_0(x) \right| \tag{43}$$

Эта функция называется ***статистикой Колмогорова*** и представляет собой максимальное отклонение эмпирической функции распределения $\hat{F}_n(x)$ от гипотетической (теоретической) функции распределения $F(x)$.

Колмогоров доказал, что при $n \to \infty$ закон распределения случайной величины $\sqrt{n} \cdot D_n$ стремится к закону распределения Колмогорова независимо от вида распределения случайной величины X, то есть

$$P\left(\sqrt{n} \cdot D_n < x\right) \to K(x) \tag{44}$$

где $K(x)$– функция распределения Колмогорова. Для нее составлена таблица значений, которой можно пользоваться при $n \geq 20$:

α	0,1	0,05	0,02	0,01	0,001
x_0	1,224	1,358	1,520	1,627	1,950

Найдем такое значение D_0, при котором выполняется равенство

$$P(D_n > D_0) = \alpha \tag{45}$$

Рассмотрим уравнение

$$K(x) = 1 - \alpha \tag{46}$$

С помощью функции распределения Колмогорова найдем корень x_0 этого уравнения, тогда

$$P\left(\sqrt{n} \cdot D_n < x_0\right) = 1 - \alpha \tag{47}$$

Следовательно, вероятность

$$P\left(\sqrt{n} \cdot D_n > x_0\right) = \alpha \tag{48}$$

Таким образом,

$$D_0 = {x_0}/{\sqrt{n}} \tag{49}$$

Если $D_n < D_0$, то гипотеза H_0 принимается, в противном случае гипотеза H_0 отвергается.