Теория вероятностей и математическая статистика (конспект лекций)

Лектор: Власов Павел Александрович, pvlx@mail.ru
МГТУ им. Н. Э. Баумана, Москва, осенний семестр, 2018-2019

$$\exists\forall \in\ni\subset\subseteq \inf \infty \cap \cup \vee \wedge\nabla \int \iint \iiint\limits_{\max diamD\to 0} \Rightarrow f(x,y)dxdy \perp \sum \prod \leqslant\geqslant \oint \mathbb{R}\nexists\{|\pi\Phi\Theta$$

Данный файл был скомпилирован 1 июля 2019 г., 14:35

Осенний семестр
Лекция №1, 04.09.2018

1 Кратные интегралы и случайные события

1.1 Двойной интеграл

1.1.1 Площадь плоской фигуры

Пусть D — некоторая область на плоскости. Если D является прямоугольником, треугольником, или, более общо, многоугольником, то понятие площади области D ввести легко. Для прямоугольника это произведение ширины на высоту, для треугольника тоже понятно как (можно школу вспомнить), многоугольник же можно разбить на несколько треугольников и определить его площадь как сумму площадей этих треугольников.

Как ввести понятие площади для произвольной области D?

Давайте поступим следующим образом:

Во-первых, для данной области D рассмотрим множество всех многоугольников M, которые целиком содержит область D.

(см. рисунок 1)

Обозначим $S^* = \inf\limits_{M} S(M)$, где $S(M)$ — площадь многоугольника M.

Давайте рассмотрим множество всех многоугольников m, которые целиком содержатся в D.

Допустим, один из таких многоугольников m выглядит как-то вот так:

(см. рисунок 2)

Обозначим $S_* = \sup\limits_m S(m)$.

Определение. *Область D на плоскости называется квадрируемой, если для неё существуют конечные значения S^*, S_* и эти значения совпадают. При этом величина $S = S^* = S_*$ называется площадью квадрируемой области D.*

По определению, множество D точек плоскости имеет площадь нуль, если D можно заключить в многоугольную фигуру сколь угодно малой площади (т. е. $\forall \varepsilon > 0$ существует M — многоугольник такой, что $D \subseteq M$, $S(M) \leqslant \varepsilon$).

Примеры:

1. Точка на плоскости имеет площадь ноль;
2. Отрезок — отрезок на плоскости можно заключить в сколь угодно малый многоугольник;
3. Гладкая кривая.

Можно доказать следующее утверждение:

Утверждение. *Пусть D — замкнутая область на плоскости. Тогда эта область квадрируемая тогда и только тогда, когда её граница имеет площадь нуль.*

Доказывать мы это утверждение не станем.

Утверждение. *Пусть L — плоская спрямляемая[1] кривая. Тогда L имеет площадь нуль.*

Утверждение приводится без доказательства.

Следствие. *Пусть D — область на плоскости, ограниченная набором спрямляемых кривых. Тогда область D — квадрируема.*

Доказательства тоже нет, хотя доказывать тут нечего (см. два предыдущих утверждения).

Замечание. *В дальнейшем мы будем рассматривать только квадрируемые области.*

[1] Спрямляемая означает, что L имеет конечную длину. — Прим. лект.

1.1.2 Задачи, приводящие к понятию двойного интеграла

1.1.2.1 Задача об объёме цилиндрического тела

Пусть

1. D — область на плоскости Oxy (замкнутая и ограниченная);
2. f — функция на плоскости, принимающая неотрицательные значения: $f\colon D \to \mathbb{R}$;
3. $f(x,\, y) \geqslant 0,\ (x,\, y) \in D$;
4. Тело G ограничено
   (a) снизу — областью D;
   (b) сверху — графиком функции $z = f(x,\, y)$;
   (c) соответствующими вертикальным прямыми, проходящими через границу области D.

Другими словами

$$G = \{\, (x,\, y,\, z) : (x,\, y) \in D,\ 0 \leqslant z \leqslant f(x,\, y)\, \}$$

(см. рисунок 3)

Задача. *Найти объём $V(G)$ тела G.*

Разобьём область D на части $D_i,\ i = \overline{1;\, n}$, так, чтобы

1. $D = \bigcup\limits_{i=1}^{n} D_i$;
2. $int^2\, D_i \cap int\, D_j = \emptyset$ при $i \neq j$.

Где $int\, M$ - множество внутренних точек множества M.

В пределах каждой подобласти D_i выберем точку $M_i,\ i = \overline{1;\, n}$.

Объём той части тела G, которая располагается над подобластью D_i, равен $\triangle V_i \approx f(M_i)\triangle S_i$, где $\triangle S_i$ — площадь области D_i. Тогда объём всего тела G

$$V(G) = \sum_{i=1}^{n} \triangle V_i \approx \sum_{i=1}^{n} f(M_i)\triangle S_i$$

[2]Уточнение от лектора: int — это interior. Не internal. — Прим. ред.

.

Полученная формула тем точнее, чем меньше размеры подобластей D_i, поэтому естественно перейти к пределу

$$V(G) = \lim_{\max diam D_i \to 0} \sum_{i=1}^{n} f(M_i) \triangle S_i$$

Замечание. *Диаметром множества M называется число $diam\, M = \sup\limits_{P,\, Q \in M} |\overrightarrow{PQ}|$*

(см. рисунок 4)

1.1.2.2 Задача о вычислении массы пластины

Пусть

1. Пластина занимает плоскую область D на плоскости Oxy;

2. $f(x,\, y)$ – значение поверхностной плотности материала пластины в точке $(x,\, y)$.

Нужно найти массу этой пластины. Давайте поступим так же, как и с цилиндрическим телом, когда искали его объём.

Разобьём область D на подобласти $D_i,\ i = \overline{1;\, n}$, так, чтобы выполнялись условия

1. $D = \bigcup\limits_{i=1}^{n} D_i$;

2. $int\, D_i \cap int\, D_j = \emptyset$ при $i \neq j$.

Тогда, считая, что размеры подобласти D_i достаточно малы, можно считать, что, в пределах этой подобласти, плотность $f(x,\, y)$ изменяется незначительно. Поэтому масса той части пластины как раз занимает подобласть D_i:

$$\triangle m_i \approx f(M_i) \triangle S_i$$

где $M_i \in D_i$ — произвольная точка, $\triangle S_i = S(D_i)$.

С учётом этого масса всей пластины

$$m = \sum_{i=1}^{n} \triangle m_i \approx \sum_{i=1}^{n} f(M_i) \triangle S_i$$

Эта функция тем точнее, чем меньше размеры D_i, поэтому

$$m = \lim_{\max diam D_i \to 0} \sum_{i=1}^{n} f(M_i) \triangle S_i,\ i = \overline{1;\, n}$$

1.1.3 Определение двойного интеграла и его свойства

Пусть D — квадрируемая замкнутая область на плоскости Oxy.

Определение. *Разбиением области D называется набор $T = \{D_1, \ldots, D_n\}$, где*

1. $D_i \subseteq D,\ i = \overline{1;\, n}$;
2. $D = \bigcup\limits_{i=1}^{n} D_i$;
3. $int\, D_i \cap int\, D_j = \emptyset$ *при* $i \neq j$.

Определение. *Диаметром разбиения T называется число $d(T) = \max\limits_{i=\overline{1;\,n}} diam\, D_i$.*

Определение. *Пусть $f\colon D \to \mathbb{R}$ — функция двух переменных.*
Двойным интегралом функции f по области D называется число

$$\iint\limits_D f(x,\, y)\, dx\, dy = \lim_{d(T)\to 0} \sum_{i=1}^{n} f(M_i)\triangle S_i$$

,

где $M_i \in D_i,\ i = \overline{1;\, n};\ \triangle S_i = S(D_i),\ i = \overline{1;\, n};\ T = \{D_1, \ldots, D_n\}$.

Замечание. *В определении подразумевается, что указанный предел существует, конечен и не зависит от выбора разбиения T области D и способа выбора точек M_i.*

Определение. *Функция f, для которой существует $\iint\limits_D f\, dx\, dy$, называется интегрируемой в D.*

1.1.3.1 Свойства двойного интеграла

1. Если область D имеет площадь S, то $\iint\limits_D 1 \cdot dx\, dy = S$.

2. **Линейность.**

   (a) Если f и g интегрируемы в D, то $f \pm g$ также интегрируема в D, причём

   $$\iint\limits_D (f \pm g)\, dx\, dy = \iint\limits_D f\, dx\, dy \pm \iint\limits_D g\, dx\, dy$$

   (b) Если f интегрируема в D, то $c \cdot f$, где $c = const$, интегрируема в D, и

   $$\iint\limits_D c \cdot f\, dx\, dy = c \iint\limits_D f\, dx\, dy$$

3. **Аддитивность.**

   Пусть

   (a) $D = D_1 \cup D_2$,
   (b) $int\, D_i \cap int\, D_2 = \emptyset$,
   (c) f интегрируема в D_1,
   (d) f интегрируема в D_2.

   Тогда f интегрируема в D, причём

$$\iint\limits_{D} f\,dx\,dy = \iint\limits_{D_1} f\,dx\,dy + \iint\limits_{D_2} f\,dx\,dy$$

4. Пусть

   (a) $f(x,\,y) \geqslant 0$ (в D),
   (b) f интегрируема (в D).

   Тогда $\iint\limits_{D} f\,dx\,dy \geqslant 0$ *(аналогично для* $\leqslant$*)*.

5. Если

   (a) $f_1(x,\,y) \geqslant f_2(x,\,y)$ (в D),
   (b) f_1 и f_2 интегрируемы (в D),

   то

$$\iint\limits_{D} f_1\,dx\,dy \geqslant \iint\limits_{D} f_2\,dx\,dy$$

6. **Теорема об оценке модуля двойного интеграла.**

   Пусть f интегрируема в D. Тогда $|f|$ также интегрируема в D, причём

$$\left|\iint\limits_{D} f\,dx\,dy\right| \leqslant \iint\limits_{D} |f|\,dx\,dy$$

   .

7. **Теорема об оценке двойного интеграла.**

   Пусть

   (a) $f,\,g$ интегрируемы (в D);
   (b) $m \leqslant f(x,\,y) \leqslant M$ (в D);

(c) $g(x, y) \geqslant 0$ (в D).

Тогда

$$m \iint\limits_D g(x, y)\, dx\, dy \leqslant \iint\limits_D f(x, y)\, g(x, y)\, dx\, dy \leqslant M \iint\limits_D g(x, y)\, dx\, dy$$

Следствие. *Если $g(x, y) \equiv 1$ в D, то свойство 7 принимает вид*

$$m \cdot S \leqslant \iint\limits_D f(x, y)\, dx\, dy \leqslant M \cdot S$$

где S — площадь области D.

8. **Теорема о среднем значении для двойного интеграла.**

Пусть

(a) f — непрерывна в D;

(b) D — квадрируемая линейно связная замкнутая область.

Тогда $\exists M_0 \in D$ такая, что

$$f(M_0) = \frac{1}{S} \iint\limits_D f(x, y)\, dx\, dy \tag{1}$$

где S — площадь области D.

Замечание. *Величину в правой части формулы 1 называют средним значением функции f в области D.*

9. **Обобщённая теорема о среднем значении.**

Пусть

(a) f непрерывна (в D);

(b) g интегрируема (в D);

(c) g знакопостоянна (опять в D);

(d) D — линейно связная замкнутая область.

Тогда $\exists M_0 \in D$ такая, что

$$\iint\limits_D f(x, y)\, g(x, y)\, dx\, dy = f(M_0) \iint\limits_D g(x, y)\, dx\, dy$$

Замечание. *Свойство 8 является следствием свойства 9, для* $g(x, y) = 1$.

Первая лекция — 102 человека. Делаем ставки, господа!

Лекция №2, 11.09.2018

1.1.4 Вычисление двойного интеграла

Определение. *Повторным интегралом называется выражение*

$$\int\limits_a^b dx \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy$$

Значением повторного интеграла называется число

$$\int\limits_a^b F(x)\, dx$$

где

$$F(x) = \int\limits_{\varphi_1(x)}^{\varphi_2(2)} f(x, y)\, dy,\ x \in [a, b]$$

Определение. *Область D на плоскости Oxy называется y-правильной, если её можно задать в следующем виде:*

$$D = \{(x, y) \colon a \leqslant x \leqslant b,\ \varphi_1(x) \leqslant y \leqslant \varphi_2(x)\}$$

(см. рисунок 5)

Замечание. *Если D — y-правильная область, то любая прямая, параллельная оси Oy, либо пересекает границу область D не более чем в двух точках. либо содержит участок границы целиком.*

x-правильная область определяется аналогично.

Определение. *Область D называется x-правильной, если её можно задать в следующем виде:*

$$D = \{(x,\, y)\colon c \leqslant y \leqslant d,\ \psi_1(y) \leqslant x \leqslant \psi_2(y)\}$$

Здесь мог бы быть ваш рисунок для x-правильной области,
но его украла лень редактора.

Замечание. *Любая прямая, параллельная оси Ox, пересекает границу x-правильной области либо не более чем в двух точках, либо содержит участок границы целиком.*

Теорема. *Пусть*

1. $\exists \iint\limits_D f(x,\, y)\, dx\, dy = I$
2. *Область D является y-правильной и задаётся в виде*

$$D = \{(x,\, y) : a \leqslant x \leqslant b,\ \varphi_1(x) \leqslant y \leqslant \varphi_2(x)\}$$

3. $\forall x \in [a, b]\ \exists \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x,\, y)\, dy \overset{\text{обозначим}}{=} F(x)$

Тогда

1. *Существует повторный интеграл* $\int\limits_a^b dx \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x,\, y)\, dy = \int\limits_a^b F(x) dx \overset{\text{обозначим}}{=} I_{\text{повт.}}$
2. $I_{\text{повт.}} = I$

Доказательство. *Без доказательства.*

Замечание. *Если область D является x-правильной, т. е. задаётся в виде*

$$D = \{(x,\, y)\colon c \leqslant y \leqslant d,\ \psi_1(x) \leqslant y \leqslant \psi_2(x)\}$$

при этом $\forall y \in [c, d]$ существует интеграл

$$\int\limits_{\psi_1(y)}^{\psi_2(y)} f(x,\, y)\, dx = G(y)$$

,

что

$$\iint\limits_D f(x,\, y)\, dx\, dy = \int\limits_c^d G(y)\, dy = \int\limits_c^d dy \int\limits_{\psi_1(y)}^{\psi_2(y)} f(x,\, y)\, dx$$

Если область D не является правильной в направлении какой-то оси, то следует разбить её на правильные подобласти и воспользоваться свойством аддитивности интеграла.

Пример. *D не является правильной в направлении Oy.*

(см рисунок 6)

Но $D = D_1 \cup D_2 \cup D_3$. Каждая из D_i является y-правильной.
Интеграл можно высчитать как:

$$\iint\limits_D f\,dx\,dy = \iint\limits_{D_1} f\,dx\,dy + \iint\limits_{D_2} f\,dx\,dy + \iint\limits_{D_3} f\,dx\,dy$$

1.1.5 Замена переменных в двойном интеграле

Пусть

1. $I = \iint\limits_D f(x,\,y)\,dx\,dy$
2. Есть область D_{xy} очень сложной формы *(см. рисунок 7)*.

Предположим, что мы подобрали

1. Область D_{uv} более простой формы *(см. рисунок 8)*;
2. Отображение Φ: $D_{uv} \to D_{xy}$

$$\Phi = \begin{cases} x = x(u,\,v) \\ y = y(u,\,v) \end{cases}$$

Тогда вычисление двойного интеграла I можно упростить.

Теорема. ***О замене переменных в двойном интеграле.***

1. *$D_{xy} = \Phi(D_{uv})$;*
2. *Φ непрерывно*[3] *и непрерывно дифференцируемо*[4]*;*
3. *Φ биективно;*

[3] Т. е. $x(u,v)$, $y(u,v)$ непрерывны. — Прим. лект.
[4] x'_u, x'_v, y'_u, y'_v существуют и непрерывны. — Прим. лект.

4. *Якобиан отображения Φ не равен нулю в D_{uv}:*

$$J_\Phi = \begin{vmatrix} x_u' & x_v' \\ y_u' & y_v' \end{vmatrix} \neq 0$$

5. *f интегрируема в D_{xy}.*

Тогда

$$\iint\limits_{D_{xy}} f(x,\,y)\,dx\,dy = \iint\limits_{D_{uv}} f(x(u,\,v),\,y(u,\,v))\,|J_\Phi(u,\,v)|\,du\,dv$$

Доказательство. *Без доказательства.*

Замечание. *Сформулированная теорема останется справедливой в том случае, если условия 2, 3, 4 нарушаются в отдельных точках или вдоль конечного числа прямых площади 0.*

Замечание. *Можно провести аналогию с «обычным» определённым интегралом. При замене*

$$x = g(t)$$

получаем

$$\int\limits_a^b f(x)dx = \int\limits_{\underline{g^{-1}(a)}}^{\underline{g^{-1}(b)}} f(\underline{g(t)})\,\underline{g'(t)}\,dt$$

В нашем случае, при замене

$$\Phi = \begin{cases} x = x(u,\,v) \\ y = y(u,\,v) \end{cases}$$

получается

$$\iint\limits_{D_{xy}} f(x,\,y)\,dx\,dy = \iint\limits_{\underline{\Phi^{-1}_{D_{xy}}}} f(\underline{x(u,\,v)},\,\underline{y(u,\,v)})\,\underline{|J_\Phi(u,\,v)|}\,du\,dv$$

1.1.5.1 Использование полярной системы координат как пример замены переменных

(см. рисунок 9)

Полярная система координат состоит из точки-полюса и полярного луча. Положение определяется как $(\rho,\, \varphi)$.

$\rho = |\overrightarrow{OM}|$ — полярный радиус, $\rho \geqslant 0$.

φ — полярный угол; φ равен углу, на который нужно повернуть полярный луч OP против часовой стрелки так, чтобы его направление совпало с направлением радиус-вектора точки M.

Это был такой экскурс на первый курс. Теперь — зачем это нужно нам с вами.

Переходим в полярную систему координат, заменив переменные

$$x = \rho \cos \varphi$$
$$y = \rho \sin \varphi$$

Получим

$$I = \iint\limits_{D_{xy}} f(x,\, y)\, dx\, dy = \iint\limits_{D_{\rho\varphi}} f(\rho \cos \varphi,\, \rho \sin \varphi)\, J_\Phi\, d\rho\, d\varphi$$

.

Таким образом, приводя к используемым ранее обозначениям: $u \Leftrightarrow \rho$, $v \Leftrightarrow \varphi$,

$$J_\Phi = \begin{vmatrix} x'_u & x'_v \\ y'_u & y'_v \end{vmatrix} = \begin{vmatrix} \cos \varphi & -\rho \sin \varphi \\ \sin \varphi & \rho \cos \varphi \end{vmatrix} = \rho \overbrace{\begin{vmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{vmatrix}}^{\text{(равен единице)}} = \rho$$

В итоге:

$$I = \iint\limits_{D_{xy}} f(x,\, y)\, dx\, dy = \iint\limits_{D_{\rho\varphi}} f(\rho \cos \varphi, \rho \sin \varphi)\, \rho\, d\rho\, d\varphi$$

1.1.6 Приложения двойного интеграла

1.1.6.1 Вычисление площади плоской фигуры

Пусть фигура занимает область D на плоскости Oxy. Тогда площадь этой фигуры равна двойному интегралу

$$S(D) = \iint\limits_D 1 \cdot dx\, dy$$

(см. свойство 1 двойного интеграла: если область D имеет площадь S, то $\iint\limits_D 1 \cdot dx\, dy = S$)

1.1.6.2 Вычисление массы пластины

Пусть

1. Пластина занимает область D на плоскости Oxy;
2. $\mu(x,\, y)$ — значение поверхностной плоскости материала пластины в точке (x, y).

Тогда масса этой пластины

$$m = \iint\limits_{D} \mu(x,\, y)\, dx\, dy$$

(см. задачу о вычислении массы пластины)

1.1.6.3 Вычисление объёма цилиндрического тела

Пусть

1. G — тело в пространстве $Oxyz$;
2. $G = \{(x,\, y,\, z)\colon (x,\, y) \in D_{xy},\; z_1(x,\, y) \leqslant z \leqslant z_2(x,\, y)\}$.

(см. рисунок 10)

Тогда объём тела G можно найти по следующей формуле:

$$V(G) = \iint\limits_{D_{xy}} [z_2(x,\, y) - z_1(x,\, y)]\, dx\, dy \tag{2}$$

Замечание. *В ранее рассмотренной задаче о вычислении объёма цилиндрического тела тело ограничивалось $f(x,\, y) \geqslant 0$ и плоскостью Oxy. Объём такого тела равен*

$$V(G) = \lim_{\max diam D_i \to 0} \sum_{i=1}^{n} f(M_i) \triangle S_i$$

Формула 2 является обобщением этого старого результата.

1.2 Тройной интеграл

1.2.1 Объём тела

Определение. *Телом будем называть замкнутую ограниченную область пространства.*

Как ввести понятие объёма тела?

Если тело G является кубом, или, более общо, многогранником, то понятие объёма можно ввести элементарным образом. А как быть, когда G — тело произвольной формы?

Поступим примерно так же, как мы поступали на прошлой лекции. Давайте рассмотрим множество всех многогранников m, целиком содержащихся в G. Пусть $V_* = \sup\limits_m V(m)$. Теперь рассмотрим множество всех многогранником M, целиком содержащихся в G. Обозначим $V^* = \inf\limits_M V(M)$.

Определение. *Тело G называется кубируемым, если существуют конечные значения V_*, V^*, причём V_*, V^*. При этом $V = V_* = V^*$ называется объёмом кубируемого тела G.*

Определение. *Говорят, что множество G точек пространства имеет объём нуль, если G можно заключить в многогранник сколь угодно малого объёма, т. е. $\forall \varepsilon > 0$ $\exists Q$ — многогранник такой, что $G \leqslant Q$.*

Замечание. *Можно показать, что точка, гладкая кривая, гладкая поверхность в пространстве имеют объём нуль.*

Теорема. *Пусть G — тело. Тогда G кубируемо тогда и только тогда, когда границы G имеют объём нуль.*

Доказательство. *Без доказательства.*

В дальнейшем мы будем рассматривать только кубируемые тела.

Лекция №3, 18.09.2018

1.2.2 Определение тройного интеграла

Давайте сперва рассмотрим задачу о вычислении массы тела.

Задача. *Пусть тело занимает область G в пространстве $Oxyz$. $\mu(x, y, z)$ — значение плотности материала этого тела в точке с координатами (x, y, z). Требуется найти массу тела G.*

Здесь мог бы быть ваш рисунок для задачи о вычислении массы тела, но его украла лень редактора.

Разобьём тело G на непересекающиеся части G_i, а точнее

$$G = \bigcup_{i=1}^{n} G_i,\ int\, G_i \cap int\, G_j = \emptyset,\ \ i \neq j$$

В пределах каждой области G_i выберем точку $M_i \in G_i$, $i = \overline{1;\, n}$. Считая, что размеры части G_i достаточно малы, можно полагать, что функция плотности μ не очень сильно изменяется в пределах области G_i, поэтому $\mu(x,\, y,\, z) \approx f(M_i)$, $(x,\, y,\, z) \in G_i$.

Тогда масса части G_i

$$m(G_i) \approx \mu(M_i) \triangle V_i$$

где $\triangle V_i$ — объём части G_i.

Тогда масса тела G

$$m(G) = \sum_{i=1}^{n} m(G_i) \approx \sum_{i=1}^{n} \mu(M_i) \triangle V_i$$

Эта функция тем точнее, чем меньше размеры G_i, поэтому естественно перейти к пределу

$$m(G) = \lim_{\max\limits_{i=\overline{1;\, n}} diam\, G_i \to 0} \sum_{i=1}^{n} \mu(M_i) \triangle V_i$$

Пусть G — тело в пространстве $Oxyz$, определена

$$f : G \to \mathbb{R}$$

— числовая функция.

Разобьём тело G на части G_i, $i = \overline{1;\, n}$ так, как это было сделано в задаче о вычислении массы тела.

Обозначим $T = \{G_1,\, \ldots,\, G_n\}$ — разбиение тела G.

Определение. *Тройным интегралом функции f по области G называется число*

$$\iiint\limits_G f(x,\, y,\, z)\, dx\, dy\, dz = \lim_{d(T) \to 0} \sum_{i=1}^{n} f(M_i) \triangle V_i$$

Замечание. *Если указанный в определении предел существует и конечен, то функция f называется интегрируемой в области G.*

Замечание. *В дальнейшем для сокращения обозначений будем иногда вместо*

$$\iiint\limits_G f(x,\, y,\, z)\, dx\, dy\, dz$$

писать

$$\iiint\limits_G f\, dV$$

1.2.2.1 Свойства тройного интеграла

Свойства тройного интеграла полностью аналогичны свойствам 1-9 двойного интеграла. В записи этих свойств вместо интеграла по области D $\iint\limits_D f(x,\, y)\, dx\, dy$ будет $\iiint f(x,\, y,\, z)\, dx\, dy\, dz$. Т. е. можно производить преобразования $D \Rightarrow G$, $\iint \Rightarrow \iiint$, $(x,\, y) \Rightarrow (x,\, y,\, z)$, $S(D) \Rightarrow V(G)$.

Свойство 1 будет при этом формулироваться так:

$$\iiint\limits_G 1 \cdot dx\, dy\, dz = V(G)$$

Замечание. *Механический смысл тройного интеграла: если $f(x,\, y,\, z) \geqslant 0$ — плотность материала в G, то $\iiint\limits_G f\, dV = m(G)$.*

1.2.3 Вычисление тройного интеграла

Пусть G — тело в пространстве $Oxyz$.

Определение. *Тело G называется z-правильным, если его можно задать в следующем виде:*

$$G = \{(x,\, y,\, z)\colon\, (x,\, y) \in D_{xy},\ z_1(x,\, y) \leqslant z \leqslant z_2(x,\, y)\} \tag{3}$$

где D_{xy} — область на плоскости Oxy.

(см. рисунок 11)

Замечание. *Любая прямая, параллельная оси z, пересекает границу z-правильного тела не более в двух точках, либо содержит участок границы целиком.*

Зачем нам всё это нужно? Вот зачем:

Теорема. *Пусть*

1. $\exists \iiint\limits_G f(x,\, y,\, z)\, dx\, dy\, dz = I$;

2. *Область G является z-правильной и задаётся формулой 3;*

3. *Для каждой фиксированной точки $(x,\, y) \in D_{xy}$ существует интеграл*

$$\int\limits_{z_1(x,y)}^{z_2(x,y)} f(x,\, y,\, z)\, dz \overset{\textit{обозначим}}{=} F(x,\, y)$$

Тогда

1. *Существует повторный интеграл*

$$\iint\limits_{D_{xy}} F(x,\, y)\, dx\, dy = \iint\limits_{D_{xy}} dx\, dy \int\limits_{z_1(x,y)}^{z_2(x,y)} f(x,\, y,\, z)\, dz \overset{\textit{обозначим}}{=} I_{\textit{повт.}}$$

2. $I = I_{\textit{повт.}}$

Замечание. *Для z-правильной области, заданной формулой 3, справедлива формула*

$$\iiint\limits_G f(x,\, y,\, z)\, dx\, dy\, dz = \iint\limits_{D_{xy}} dx\, dy \int\limits_{z_1(x,y)}^{z_2(x,y)} f(x,\, y)\, dz$$

Замечание. *Если при этом область D_{xy} является y-правильной и задаётся в следующем виде*[5]

$$D = \{(x,\, y)\colon a \leqslant x \leqslant b,\ \varphi_1(x) \leqslant y \leqslant \varphi_2(x)\}$$

то

$$\iiint\limits_G f(x,\, y,\, z)\, dx\, dy\, dz = \int\limits_a^b dx \int\limits_{\varphi_1(x)}^{\varphi_2(x)} dy \int\limits_{z_1(x,y)}^{z_2(x,y)} f(x,\, y,\, z)\, dz$$

1.2.4 Замена переменных в тройном интеграле

Теорема. *Пусть*

1. $G_{xyz} = \Phi(G_{uv\omega})$, *где* $\Phi = \begin{cases} x = x(u,\, v,\, \omega) \\ y = y(u,\, v,\, \omega) \\ z = z(u,\, v,\, \omega) \end{cases}$

[5]Стандартное определение y-правильной области. — Прим. ред.

2. *Φ биективно;*

3. *Φ непрерывно и непрерывно дифференцируемо в $G_{uv\omega}$;*

4. *Якобиан*

$$J_\Phi = \begin{vmatrix} x'_u & x'_v & x'_\omega \\ y'_u & y'_v & y'_\omega \\ z'_u & z'_v & z'_\omega \end{vmatrix} \neq 0 \ (\text{в } G_{uv\omega})$$

.

Тогда

$$\iiint\limits_{G_{xyz}} f(x,\, y,\, z)\, dx\, dy\, dz = \iiint\limits_{G_{uv\omega}} f(x(u,\, v,\, \omega), y(u,\, v,\, \omega), z(u,\, v,\, \omega))|J_\Phi|\, du\, dv\, d\omega$$

Доказательство. *Без доказательства.*

1.2.4.1 Пример. Цилиндрическая система координат

В цилиндрической системе координат положение задаётся как (ρ, φ, z). ρ и φ имеют тот же смысл, что и полярной системе координат, а z имеет тот же смысл, что и в декартовой системе координат.

(см. рисунок 12)

Связь цилиндрической системы координат с декартовой:

$$x = \rho\cos\varphi$$
$$y = \rho\sin\varphi$$
$$z = z$$

$$J_{\text{цил.}} = \begin{vmatrix} \cos\varphi & -\rho\sin\varphi & 0 \\ \sin\varphi & \rho\cos\varphi & 0 \\ 0 & 0 & 1 \end{vmatrix} = \rho$$

Получается

$$\iiint\limits_{G_{xyz}} f(x,\, y,\, z)\, dx\, dy\, dz = \iiint\limits_{G_{\rho\varphi z}} f(\rho\cos\varphi, \rho\sin\varphi, z)\rho\, d\rho\, d\varphi\, dz$$

1.2.4.2 Пример. Сферическая система координат

Положение в сферической системе координат задаётся при помощи (ρ, φ, θ).

(см. рисунок 13)

Действуют следующие ограничения: $\rho \geqslant 0$, $\varphi \in [0,\, 2\pi)$, $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.
Связь сферической системы координат с декартовой системой координат:

$$\begin{aligned} x &= \rho\cos\theta\cos\varphi \\ y &= \rho\cos\theta\sin\varphi \\ z &= \rho\sin\theta \end{aligned}$$

Якобиан[6]

$$J_{\text{сф.}} = \rho^2\cos\theta$$

Таким образом,

$$\iiint\limits_{G_{xyz}} f(x,\, y,\, z)\, dx\, dy\, dz = \iiint\limits_{G_{\rho\varphi\theta}} f(\rho\cos\theta\cos\varphi,\, \rho\cos\theta\sin\varphi,\, z\sin\theta)\, \rho^2\cos\theta\, d\rho\, d\varphi\, d\theta$$

...[7]

2 Теория вероятностей

2.1 Случайные события

2.1.1 Основные понятия

2.1.1.1 Случайные эксперименты

Определение. *Случайным называют эксперимент, результат которого невозможно точно предсказать заранее.*

Пример. *Бросают монетку. Возможные исходы: выпадение герба или решки.*
$\Omega = \{$ *Герб* , *Решка* $\}$ — *множество элементарных исходов.* $|\Omega| = 2$

Пример. *Бросают игральную кость:* $\Omega = \{$ *“1”, “2”, “3”, “4”, “5”, “6”*$\}$, $|\Omega| = 6$

[6] Вывод якобиана можно проверить самостоятельно. — Прим. лект.

[7] Запись и изучение теории, связанной с четвертными интегралами, пятикратными интегралами, шестикратными интегралами и т. д., а также обобщение всех определений и теорем на случай n-кратного интеграла предлагается сделать самостоятельно. — Прим. ред.

Пример. *Из колоды в 36 карт последовательно извлекают 2 карты (без возвращения). Исход можно описать парой (x_1, x_2), где x_i — номер карты при i-ом извлечении.*

Тогда

$$\Omega = \{(x_1, x_2)\colon x_1, x_2 \in \{1, \ldots, 36\},\ x_1 \neq x_2\},\ \ |\Omega| = 36 \cdot 35$$

Пример. *Бросают монету до первого появления герба. Наблюдаемый результат — число бросков.*

$$\Omega = \{1, 2, \ldots\} = \mathbb{N},\ \ |\Omega| = \aleph_0$$

$\aleph_0$ *— одно из обозначений мощности счётного множества*[8].

Пример. *Происходит выстрел по плоской мишени. Наблюдаемый результат: пара (x, y) — координаты точки попадания пули. $\Omega = \mathbb{R}^2$*

Определение. *Множество Ω всех исходов данного случайного эксперимента называют пространством элементарных исходов.*

Замечание. *Всюду в дальнейшем будем предполагать, что*

1. *Каждый элементарный исход является «неделимым», т. е. он не может быть разбит на более «мелкие» исходы;*

2. *В результате каждого эксперимента обязательно имеет место ровно один из входящих в Ω элементарных исходов.*

Нестрогое определение. *Событием будем называть произвольное подмножество множества элементарных исходов Ω.*

Определение. *Говорят, что в результате случайного эксперимента наступило событие A, если в результате данного эксперимента был реализован один из входящих в A элементарных исходов.*

Пример. *Из колоды в 36 карт извлекают одну карту.*

$$\Omega = \{6_{\textit{пик}}, \ldots, T_{\textit{пик}}, 6_{\textit{треф}}, \ldots, \ldots, T_{\textit{червей}}\},\ \ |\Omega| = 36$$

Можно определить событие $A = \{\text{извлечена карта красной масти}\}$, т. е. $A = \{6_{\textit{бубей}}, \ldots, T_{\textit{бубей}}, 6_{\textit{червей}}, \ldots, T_{\textit{червей}}\}$, $|A| = 18$. Если в результате эксперимента извлечена $6_{\textit{бубей}}$, то всё событие A целиком «наступило»[9].

Определение. *Событие A называется следствием события B, если из того, что произошло B, следует то, что произошло A, т. е. $B \subseteq A$.*

[8] См. статьи Википедии «Счётное множество» и «Иерархия алефов». В любом случае, сильно сомневаюсь, что это потребуется на РК или экзамене. — Прим. ред.

[9] Эта формулировка очень грубая. — Прим. лект.

Замечание. *Любое множество Ω содержит в себе два подмножества: Ω и $\emptyset$. События, соответствующие данным множествам, называются невозможным и достоверным соответственно. Эти события называются несобственными событиями. Все остальные события называются собственными.*

Пример. *Из урны, содержащей два красных и три синих шара, извлекают один шар. Возможные события: $A = \{$ извлечённый шар является красным или синим $\}$ — является достоверным, $B = \{$ извлечён белый шар $\}$ — невозможным.*

Лекция №4, 25.09.2018

2.1.2 Операции над событиями

События — множества элементарных исходов. Следовательно, над ними можно выполнять все операции над множествами. При этом вводится следующая терминология:

- Объединение множеств принято называть суммой событий: $A \cup B = A + B$;
- Пересечение множеств называют произведением событий: $A \cap B = A \cdot B$;
- $A \setminus B$ называют разностью событий A и B;
- Дополнение A называют событием, противоположным A: $\overline{A} = \Omega \setminus A$.

2.1.2.1 Основные свойства операций над событиями

1. $A + B = B + A$ (коммутативность);
2. $AB = BA$ (коммутативность);
3. $(A + B) + C = A + (B + C)$ (ассоциативность);
4. $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ (ассоциативность);
5. $A \cdot (B + C) = AB + AC$ (дистрибутивность);
6. $A + (BC) = (A + B)(A + C)$ (дистрибутивность);
7. $\overline{\overline{A}} = A$;
8. $A + A = A$ (идемпотентность);
9. $A \cdot A = A$ (идемпотентность);
10. $\overline{A + B} = \overline{A}\,\overline{B}$ (законы де Моргана);

11. $\overline{AB} = \overline{A} + \overline{B}$ (законы де Моргана);

12. $A \subseteq B \Rightarrow A + B = B$;

13. $A \subseteq B \Rightarrow AB = A$;

14. $A \subseteq B = \overline{B} \subseteq \overline{A}$.

Определение. *События A и B называются несовместными, если их произведение пусто ($AB = \emptyset$). В противном случае события A и B называются совместными.*

Определение. *События $A_1, \ldots, A_n, \ldots$ называются*[10]

1. *Попарно несовместными, если $A_i \cdot A_j = \emptyset$ при $i \neq j$;*

2. *Несовместными в совокупности, если $A_1 \cdot \ldots \cdot A_n \cdot \ldots = \emptyset$.*

Замечание. *Из попарной несовместности следует несовместность в совокупности. Обратное в общем случае неверно.*

2.1.3 Классическое определение вероятности

Пусть

1. Ω — пространство исходов некоторого случайного эксперимента ($|\Omega| = N < \infty$)[11];

2. По условиям эксперимента нет оснований предпочесть тот или иной элементарный исход остальным (в таком случае говорят, что все элементарные исходы равновозможны);

3. Существует событие $A \subseteq \Omega$, мощность $|A| \overset{\text{(обозначим)}}{=} N_A$

Тогда

Определение. *Вероятностью осуществления события A называют число*

$$P\{A\} = \frac{N_A}{N}$$

Задача-пример. *Два раза бросают игральную кость. Задано событие*

$$A = \{ \text{ сумма выпавших очков больше или равна одиннадцати } \}$$

Найти $P(A)$.

[10] Многоточия в конце перечислений обязательны, т. к. обозначают, что в перечислении может быть бесконечное число величин. — Прим. лект.

[11] Запись $x < \infty$ означает, что x конечно. Напротив, запись $x \leqslant \infty$ означает, что x либо конечно, либо бесконечно. — Прим. ред.

Решение. *Определим исход:* $(x_1,\, x_2)$ *— упорядоченная пара, где* x_i *— кол-во очков, выпавших при i-ом броске,* $i = \overline{1;\, 2}$ *(*$N = |\Omega| = 36$*).*

В событии A можно выделить следующие исходы:

$$A = \{(5,\, 6), (6,\, 5),\, (6,\, 6)\}$$

Тогда $N_A = |A| = 3$*, и, в соответствии с определением, получается*

$$P(A) = \frac{N_A}{N} = \frac{3}{36} = \frac{1}{12}$$

2.1.3.1 Свойства вероятности (в соответствии с классическим определением)

1. Вероятность $P(A) \geqslant 0$ (неотрицательна).

2. $P(\Omega) = 1$.

3. Если A, B — несовместные события, то $P(A + B) = P(A) + P(B)$.

Доказательства этих свойств:

1. Т. к. $N_A \geqslant 0$, $N > 0$, то следует $P(A) = \frac{N_A}{N} \geqslant 0$.

2. Принимая во внимание, что $N_\Omega = |\Omega| = N$, получается

$$P(\Omega) = \frac{N_\Omega}{N} = \frac{N}{N} = 1$$

3. Т. к. Ω — конечно, $A, B \subseteq \Omega$, то получается, что A, B конечны. Существует формула[12]

$$|A + B| = |A| + |B| - |AB|$$

Т. к. A и B — несовместные, то $AB = \emptyset$, из чего следует, что $N_{a+b} = N_a + B_b$. Таким образом,

$$P(A + B) = \frac{N_{a+b}}{N} = \frac{N_a + N_b}{N} = \frac{N_a}{N} + \frac{N_b}{N} = P(A) + P(B)$$

Замечание. *У классического определения есть следующие недостатки:*

- *Оно неприменимо в случае бесконечного числа элементарных исходов;*

- *Оно неприменимо в случае, когда некоторые элементарные исходы являются более или менее предпочтительными.*

Частично эти недостатки исправляет геометрическое определение вероятности.

[12] Её называют формулой включений и исключений. — Прим. лект.

2.1.4 Геометрическое определение вероятности

Геометрическое определение вероятности является обобщением классического определения на случай, когда $|\Omega| = \infty$.

Пусть

1. $\Omega \subseteq \mathbb{R}^n$;

2. $\mu(\Omega) < \infty$, где μ — некая мера.

   Если $n = 1$, то μ — это длина; если $n = 2$, то μ — площадь; если $n = 3$ — объём. Можно определить меры и при больших n;

3. Возможность принадлежности некоторого элементарного исхода случайного эксперимента событию $A \subseteq Q$ пропорциональна мере этого события и не зависит от формы события A и его расположения внутри Ω.

Тогда

Определение. *Вероятностью случайного события $A \subseteq \Omega$ называют число*

$$P\{A\} = \frac{\mu(A)}{\mu(\Omega)}$$

Пример. ***Задача о встрече.***

Два человека договорились встретиться в определённом месте с 12 до 13 часов. При этом каждый из них может прийти в условленное место в любое время этого промежутка (т. е. равновероятно в начале часа, в середине и в конце).

Придя на место, каждый из них ждёт 15 минут и уходит. Какова вероятность того, что они встретятся?

Решение. *Исходы: (x_1, x_2) — упорядоченные пары, где $x_i \in [0; 1]$ — время появления в условленном месте i-ого человека (отсчитывая от 12 часов дня).*

События Ω будут состоять из всех таких элементарных исходов

$$\Omega = \{(x_1, x_2)\colon x_i \in [0; 1]\} = [0; 1] \times [0; 1]$$

Зададим событие «встреча» как

$$A = \left\{(x_1, x_2)\colon |x_1 - x_2| \leqslant \frac{1}{4}\right\} \tag{4}$$

Изобразим Ω и A в виде множеств точек на плоскости.

(см. рисунок 14)

В соответствии с геометрическим определением

$$P\{A\} = \underbrace{\frac{S(A)}{S(\Omega)}}_{S(\Omega)=1} = S(A) = 1 - S(\overline{A}) = 1 - \frac{9}{16} = \frac{7}{16}$$

Ответ: $\frac{7}{16}$

Замечание. *В выражении 4 строгость/нестрогость знака сравнения не имеет значения. В дальнейшем для обозначения этого будем использовать знак* $\preccurlyeq \in \{<, \leqslant\}$ [13]

Замечание. *Очевидно, из геометрического определения вероятности можно доказать те же свойства функции P, что и для классического определения.*

Недостаток геометрического определения заключается в том, что оно не учитывает возможность того, что некоторые области внутри Ω окажутся более предпочтительными, чем другие.

Например, в предыдущем примере, если появление встречающихся более вероятно в середине часа, то геометрическое определение не даст удовлетворительный результат.

2.1.5 Статистическое определение вероятности

Пусть

1. Некоторый случайный эксперимент произведён n раз;
2. При этом некоторое наблюдаемое в этом эксперименте событие A произошло n_A раз.

Определение. *Вероятностью осуществления события A называют эмпирический (т. е. найденный экспериментальным путём) предел:*

$$P(A) = \lim_{n\to\infty} \frac{n_a}{n}$$

Замечание. *Можно показать, что для статистического определения останутся в силе доказанные выше свойства вероятностей.*

Замечание. *У статистического определения полным-полно недостатков:*

1. *Никакой эксперимент не может быть произведён бесконечное много раз;*
2. *С точки зрения современной математики статистическое определение является архаизмом, т. к. не даёт достаточно базы для дальнейшего построения теории.*

[13] При ведении лекций вместо $\preccurlyeq$ использовался знак, похожий на $\leqslant$, где нижняя черта рисовалась пунктиром. Похожего знака нет в стандартном наборе знаков LaTeX, поэтому была выбрана замена. — Прим. ред.

2.1.6 Сигма-алгебра событий

Для строгого аксиоматического определения вероятности необходимо уточнить понятие события:

1. Данное выше определение события как произвольного подмножества множества Ω в случае бесконечного множества Ω приводит к противоречивой теории (см. парадокс Рассела);

2. Таким образом, необходимо в качестве события рассматривать не все возможные подмножества множества Ω, а лишь некоторые из них;

3. Набор подмножеств множества Ω, выбранных в качестве событий, должен обладать рядом свойств. Понятно, что если A и B — связанные со случайным экспериментом события и известно, что в результате эксперимента они произошли (или не произошли), то естественно знать, произошли ли события $A+B$, $A \cdot B$, $\overline{A}$, ...

   Эти соображения приводят к следующему определению.
   Пусть

1. Ω — пространство элементарных исходов, связанных с некоторым случайным экспериментом;

2. $\beta \neq \emptyset$ — система (набор) подмножеств в множестве Ω.

Определение. *β называется сигма-алгеброй*[14] *событий, если выполнены условия:*

1. Если $A \in \beta$, то $\overline{A} \in \beta$; [15]

2. Если $A_1, \ldots, A_n, \ldots \in \beta$, то $A_1 + \ldots + A_n + \ldots \in \beta$.

2.1.6.1 Простейшие следствия из аксиом сигма-алгебры

1. $\Omega \in \beta$;

2. $\emptyset \in \beta$;

3. Если $A_1, \ldots, A_n, \ldots \in \beta$, то $A_1 \cdot \ldots \cdot A_n \cdot \ldots \in \beta$;

4. Если A, $B \in \beta$; то $A \setminus B \in \beta$.

Доказательства этих следствий:

[14] При ведении лекций слово «сигма» иногда заменялось на букву δ (дельта — `\delta` в LaTeX). Буква «сигма» выглядит как σ. Лектор говорит, что корректнее всего словосочетание «сигма-алгебра» вообще не сокращать и писать полностью, не используя греческие буквы. Тем не менее, в Википедии и нескольких учебниках по Теории Вероятностей используется именно «σ-алгебра». — Прим. ред.

[15] Обратите внимание, что $A \subseteq \Omega$, но $A \in \beta$, т. к. элементы множества β — подмножества из Ω. — Прим. лект.

1. По определению $\beta \neq \emptyset \implies \exists A \subseteq Q\colon A \in \beta$; из определения сигма-алгебры (аксиома 1) $\exists A \in \beta \implies \overline{A} \in \beta$; тогда из второй аксиомы следует, что $\exists (A + \overline{A}) \in \beta$; т. к. $A + \overline{A} = \Omega$, то $\Omega \in \beta$.

2. Т. к. $\Omega \in \beta$ (по следствию 1), то, по аксиоме 1, $\overline{\Omega} \in \beta$, а $\overline{\Omega} = \emptyset$. Следовательно, $\emptyset \in \beta$.

3. Из существования событий $A_1, \ldots, A_n, \ldots \in \beta$ по аксиоме 1 следует, что существуют дополнения этих событий $\overline{A_1}, \ldots, \overline{A_n}, \ldots \in \beta$. По аксиоме 2 следует существование объединения $\overline{A_1} + \ldots + \overline{A_n} + \ldots \in \beta$, и из аксиомы 1 — существование дополнения этого объединения: $\overline{\overline{A_1} + \ldots + \overline{A_n} + \ldots} \in \beta$. Из этого, по законам де Моргана, получается $\overline{\overline{A_1}} \cdot \ldots \cdot \overline{\overline{A_n}} \cdot \ldots \in \beta$, что тривиально преобразуется в $A_1 \cdot \ldots \cdot A_n \cdot \ldots \in \beta$.

4. Из свойств операций над множествами можно заключить, что $A \setminus B = A \cdot \overline{B}$. По аксиоме 1, из $B \in \beta \implies \overline{B} \in \beta$. По следствию 3, $A, \overline{B} \in \beta \implies A \cdot \overline{B} \in \beta$, что, собственно, является утверждением $A \setminus B \in \beta$.

Замечание. *В дальнейшем всегда будем предполагать, что на множестве элементарных исходов задана сигма-алгебра событий. При этом событиями будем называть элементы этой сигма-алгебры и только их.*

Замечание. *Если $|\Omega| < \infty$, то в качестве β будем рассматривать (по умолчанию) множество всех подмножеств множества Ω.*

Пример. *Человека попросили выбросить одно из трёх: камень, ножницы или бумагу.*

$$\Omega = \{\,К,\ Н,\ Б\,\}$$

Тогда

$$\beta = \{\emptyset, \{К\}, \{Н\}, \{Б\}, \{К, Н\}, \{К, Б\}, \{Н, Б\}, \{К, Н, Б\}\}$$

Лекция №5, 02.10.2018

2.1.7 Аксиоматическое определение вероятности

Пусть

1. Ω — пространство элементарных исходов некоторого случайного эксперимента;

2. β — сигма-алгебра, заданная на Ω.

Определение. *Вероятностью (вероятностной мерой) называется функция*

$$P\colon \beta \to \mathbb{R}$$

обладающая следующими свойствами:

1. $\forall A \in \beta \implies P(A) \geqslant 0$ *(аксиома неотрицательности);*
2. $P(\Omega) = 1$ *(аксиома нормированности);*
3. *Если* $A_1, \ldots, A_n, \ldots$ *— попарно несовместные события, то вероятность осуществления их суммы равна сумме вероятностей осуществления каждого из них по отдельности:* $P(A_1 + \ldots + A_n + \ldots) = P(A_1) + \ldots + P(A_n) + \ldots$ *(расширенная аксиома сложения).*

Замечание 1. *Аксиомы 1-3 называются аксиомами вероятности.*

Замечание 2. *Тройка* (Ω, β, P) *называется вероятностным пространством.*

2.1.7.1 Свойства вероятностей (из аксиоматического определения)

1. $P(\overline{A}) = 1 - P(A)$;
2. $P(\emptyset) = 0$;
3. Если $A \subseteq B$, то $P(A) \leqslant P(B)$;
4. $\forall A \in \beta\colon 0 \leqslant P(A) \leqslant 1$;
5. $P(A + B) = P(A) + P(B) - P(AB)$, где $A, B \in \beta$;
6. Для любого *конечного* набора событий $A_1, \ldots, A_n$ верно

$$\begin{aligned} &P(A_1 + \ldots + A_n) = \\ &+ \sum_{1 \leqslant i_1 \leqslant n} P(A_{i_1}) \\ &- \sum_{1 \leqslant i_1 < i_2 \leqslant n} P(A_{i_1}, A_{i_2}) \\ &+ \sum_{1 \leqslant i_1 < i_2 < i_3 \leqslant n} P(A_{i_1}, A_{i_2}, A_{i_3}) - ... + ... \end{aligned}$$

Доказательства этих свойств:

1. По аксиомам 1, 2 сигма-алгебры $\exists A + \overline{A} = \Omega$; по аксиоме вероятности №2 $P(\Omega) = 1 = P(A + \overline{A})$; по аксиоме вероятности №3 (A и $\overline{A}$ несовместны), $P(A + \overline{A}) = P(A) + P(\overline{A}) = 1 \implies P(\overline{A}) = 1 - P(A)$.

2. $P(\emptyset) = P(\overline{\Omega})$; по свойству №1 $P(\emptyset) = 1 - \overbrace{P(\Omega)}^{=1 \text{ (по аксиоме 2)}} = 0$

3. $A \subseteq B \overset{\text{(по рисунку)}}{\Longrightarrow} B = A + (B \setminus A)$

(см. рисунок 15)

Тогда

$$P(B) = P(A + (B \setminus A)) = \underset{A, B\setminus A \text{ несовместны, используем аксиому 3}}{=}$$
$$= P(A) + \overbrace{P(B \setminus A)}^{\geqslant 0 \text{ по аксиоме 1}} \geqslant P(A)$$
$$\Longrightarrow P(B) \geqslant P(A)$$

4.
   (a) Неравенство $P(A) \geqslant 0$ следует из аксиомы 1.
   (b) Осталось доказать, что $P(A) \leqslant 1$.

$$\forall A \subseteq \Omega \overset{\text{по свойству 3}}{\Longrightarrow} P(A) \leqslant \overset{=1}{P(\Omega)} \Longrightarrow P(A) \leqslant 1$$

5. Для любых A, B:
   (a) $A + B = A + (B \setminus A)$,

(см. рисунок 16)

при этом $A \cdot (B \setminus A) = \emptyset$.
В соответствии с аксиомой 3,

$$P(A + B) = P(A) + P(B \setminus A) \tag{5}$$

   (b) $B = AB + (B \setminus A)$,

(см. рисунок 17)

причём $(AB)(B \setminus A) = \emptyset$.
По аксиоме 3, имеем $P(B) = P(AB) + P(B \setminus A) \implies P(B \setminus A) = P(B) - P(AB)$. Подставим результат в 5 и получим

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

6. Это свойство доказывать не станем. Оно является обобщением свойства 5 и может быть доказано из 5 с использованием метода математической индукции.

Замечание. *Иногда вместо расширенной аксиомы сложения (акс. вер. №3) рассматривают следующие две аксиомы:*

$3'$: *Для любого конечного набора попарно несовместных событий $A_1, \ldots, A_n$ вероятность осуществления их суммы равна сумме вероятностей осуществления каждого из них по отдельности.*

(аксиома сложения («обычная»);
в расширенной набор является счётным, а не конечным)

$3''$: *Для любой неубывающей последовательности событий*

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$$

и события $A = \bigcup\limits_i A_i$ верно

$$P(A) = \lim_{i \to \infty} P(A_i)$$

(аксиома непрерывности)

Замечание. *Можно доказать, что аксиома 3 эквивалентна совокупности $3'$ и $3''$.*

2.1.8 Условная вероятность

2.1.8.1 Определение условной вероятности

Пусть

1. A и B — два события, связанные с одним случайным экспериментом;

2. Дополнительно известно, что в результате эксперимента произошло событие B.

Что в этом случае можно сказать о вероятности наступления события A?

Пример. *Из колоды в 36 карт случайным образом извлекают одну карту.*

События

$$A = \{\ \textit{извлечён туз}\ \},\ B = \{\ \textit{извлечена картинка}^{16}\ \}$$

«Безусловная» (т. е. «обычная») вероятность события A: $P(A) = \frac{4}{36} = \frac{1}{9}$

«Условная» (т. е. с учётом информации о том, что произошло событие B) вероятность осуществления события A: $P_B(A) = \frac{1}{4}$

Давайте сперва дадим геометрическую интерпретацию:

Пусть $|\Omega| = N < \infty$; т. к. мы знаем, что в результате эксперимента наступило событие B, то можно рассматривать лишь те элементарные исходы, которые попали в B.

(см. рисунок 18)

Тогда событие A может осуществиться лишь в том случа, если имел место элементарный исход из AB.

Если считать все N исходов равновозможными, то

$$P_B(A) = \frac{N_{AB}}{N_B} = \frac{\left(\frac{N_{AB}}{N}\right)}{\left(\frac{N_B}{N}\right)} = \frac{P(AB)}{P(B)}$$

Определение. *Условной вероятностью осуществления события A при условии, что произошло B, называется число*[17]

$$P(A\,|\,B) = \frac{P(AB)}{P(B)},\ P(B) \neq 0$$

Замечание. *Иногда, чтобы подчеркнуть разницу, «обычную» вероятность $P(A)$ называют «безусловной».*

Если зафиксировать событие B и рассматривать $P(A\,|\,B)$ как функцию события A, то оказывается, что условная вероятность обладает всеми свойствами безусловной.

Теорема. *Пусть*

1. *Зафиксировано событие B, $P(B) \neq 0$;*
2. *$P(A\,|\,B)$ рассматривается как функция события A.*

Тогда $P(A\,|\,B)$ обладает всеми свойствами безусловной вероятности.

[16] Валет, дама, король, туз. — Прим. ред.

[17] В разговорной речи $P(A\,|\,B)$ читается как P от A при B. — Прим. лект.

Доказательство. *Докажем отдельно соответствие $P(A \mid B)$ трём аксиомам вероятности и следствиям из неё.*

1. *Докажем, что условная вероятность $P(A \mid B)$ удовлетворяет трём аксиомам вероятности:*

   (a) $P(A \mid B) = \dfrac{\overbrace{P(AB)}^{\geqslant 0}}{\underbrace{P(B)}_{>0}} \implies P(A|B) \geqslant 0.$

   (b) $P(\Omega \mid B) = \dfrac{P(\Omega B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1.$

   (c)

$$
\begin{gathered}
P(A_1 + \ldots + A_n + \ldots \mid B) = \frac{P((A_1 + \ldots + A_n + \ldots)B)}{P(B)} = \\
= \frac{1}{P(B)} \cdot P(A_1 B + A_2 B + \ldots + A_n B + \ldots) \underset{A_i, A_j \text{ несовместны, } i \neq j;\ A_i B \subseteq A_i,\ A_j B \subseteq A_j \implies (A_i B) \cap (A_j B) = \emptyset, \text{ и тогда по аксиоме вероятности №3}}{=} \\
= \frac{1}{P(B)} \cdot [P(A_1 B) + \ldots + P(A_n B) + \ldots] = \\
= \text{(ряд)} \frac{P(A_1 B)}{P(B)} + \ldots + \frac{P(A_n B)}{P(B)} + \ldots = \\
= P(A_1 \mid B) + \ldots + P(A_n \mid B) + \ldots
\end{gathered}
$$

2. *Т. к. свойства 1-6 безусловной вероятности являются прямыми следствиями из аксиом 1-3, а условная вероятность этим аксиомам удовлетворяет, то она удовлетворяет свойствам 1-6.*

При решении задач для вычисления условной вероятности используют обычно два приёма:

1. Воспользоваться формулой $P(A \mid B) = \dfrac{P(AB)}{P(B)}$

2. Можно перестроить пространство элементарных исходов и вычислять $P(A \mid B)$ как безусловную вероятность в перестроенном вероятностном пр-ве $(\Omega_1,\ \beta_1,\ P_1)$, где $\Omega_1 = B$.

Пример. *Среди 15 лотерейных билетов — пять выигрышных. Два игрока по очереди тянут по одному билету (каждый по одному разу).*

Заданы события

$$A_1 = \{ \text{ первый игрок вытащил выигрышный билет } \}$$
$$A_2 = \{ \text{ второй игрок вытащил выигрышный билет } \}$$

Нужно найти $P(A_2 \,|\, A_1)$.

Решение 1. *Определим исход* (x_1, x_2), *где* x_i *— номер билета при* i*-ом извлечении,* $i = \overline{1;\, 2}$. *Количество исходов равняется кол-ву размещений без повторений*[18] *из 15 по 2* ($N = A_{15}^2 = 15 \cdot 14$).

Тогда

1. $P(A_1) = \dfrac{N_{A_1}}{N} = \dfrac{5 \cdot 14}{15 \cdot 14} = \dfrac{1}{3}$

2. $P(A_1 A_2) \overset{A_1 A_2 \text{ — оба вытащили выигрышные билеты}}{=} \dfrac{5 \cdot 4}{15 \cdot 14} = \dfrac{2}{21}$

$$P(A_2 \,|\, A_1) = \frac{P(A_1 A_2)}{P(A_1)} = \frac{\left(\frac{2}{21}\right)}{\left(\frac{1}{3}\right)} = \frac{2}{7}$$

Решение 2. *Известно, что наступило* A_1. *Следовательно, после хода первого игрока осталось 14 билетов, из которых четыре выигрышных. Таким образом,*

$$P(A_2 \,|\, A_1) = \frac{4}{14} = \frac{2}{7}$$

Лекция №6, 09.10.2018

2.1.8.2 Формула умножения вероятностей

Теорема. ***Формула умножения вероятностей для двух событий***

Пусть

1. A, B *— события;*

2. $P(A) > 0$.

Тогда

$$P(AB) = P(A)\, P(B \,|\, A)$$

Доказательство. *Т. к.* $P(A) > 0$, *то определена условная вероятность*

$$P(B \,|\, A) = \frac{P(AB)}{P(A)}$$

из чего напрямую следует

$$P(AB) = P(A)\, P(B \,|\, A)$$

[18] Комбинаторика не преподавалась на лекциях, но был произведён краткий обзор на семинарах. См. также прил. А. — Прим. ред.

Теорема. ***Формула умножения вероятностей для n событий***
Пусть

1. $A_1, \ldots, A_n$ — *события;*
2. $P(A_1 \cdot \ldots \cdot A_{n-1}) > 0$.

Тогда

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

Доказательство.

1. *Обозначив* $k = \overline{1;\, n-1}$, *имеем* $A_1 \cdot \ldots \cdot A_k \supseteq A_1 \cdot \ldots \cdot A_{n-1}$.

   По свойству 3 вероятности $P(A_1 \cdot \ldots \cdot A_k) \geqslant P(A_1 \cdot \ldots \cdot A_{n-1}) > 0$.

   Следовательно, все условные вероятности, входящие в правую часть доказываемой формулы, определены, и можно задавать условные вероятности по типу $P(A_n \,|\, A_1 A_2 \,\ldots\, A_{n-1})$, *и, следовательно, можно пользоваться формулой умножения вероятностей для двух событий.*

2. *Последовательно применим формулу умножения вероятностей для двух событий*[19] $(P(A_{mf} B_{mf}) = P(A_{mf}) P(B_{mf} \,|\, A_{mf}))$:

$$P(\underbrace{A_1 \cdot \ldots \cdot A_{n-1}}_{A_{mf1}} \cdot \underbrace{A_n}_{B_{mf1}}) =$$

$$= P(\overbrace{\underbrace{A_1 \cdot \ldots \cdot A_{n-2}}_{A_{mf2}} \cdot \underbrace{A_{n-1}}_{B_{mf2}}}^{A_{mf1}}) \cdot P(\overbrace{A_n}^{B_{mf1}} \,|\, \overbrace{A_1 \cdot \ldots \cdot A_{n-1}}^{A_{mf1}}) =$$

$$= P(\overbrace{\underbrace{A_1 \cdot \ldots \cdot A_{n-3}}_{A_{mf3}} \cdot \underbrace{A_{n-2}}_{B_{mf3}}}^{A_{mf2}}) \cdot P(\overbrace{A_{n-1}}^{B_{mf2}} \,|\, \overbrace{A_1 \cdot \ldots \cdot A_{n-2}}^{A_{mf2}}) \cdot P(A_n \,|\, A_1 \cdot \ldots \cdot A_{n-1}) =$$

$$= \ldots =$$

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

Пример. *На семи карточках написаны буквы слова «ШОКОЛАД». Карточки тщательно перемешивают, и по очереди извлекают случайным образом три из них без возвращения первых карточек. Найти вероятность того, что эти три карточки в порядке появления образуют слово «ШОК»:*

$$A = \{ \text{ три карточки в порядке появления образуют слово «ШОК» } \}$$

[19] Т. к. обозначения A, B накладываются на уже используемые, то при иллюстрации применения этой формулы будем использовать индекс ${}_{mf}$ (multiplication formula). — Прим. ред.

Давайте введём следующие обозначения:

$$A_1 = \{\text{ на первой извлечённой карточке написано «Ш» }\}$$
$$A_2 = \{\text{ на второй извлечённой карточке написано «О» }\}$$
$$A_3 = \{\text{ на третьей извлечённой карточке написано «К» }\}$$

Тогда $A = A_1 \cdot A_2 \cdot A_3$.

$$P(A) = P(A_1A_2A_3) \overset{\text{по ф-ле умножения вероятностей}}{=} = P(A_1) \cdot P(A_2 \,|\, A_1) \cdot P(A_3 \,|\, A_1A_2)$$

Вероятность события A_1 — $\frac{1}{7}$.

Предположим, что в результате эксперимента стало доподлинно известно, что произошло событие A_1. *Тогда вероятность вытащить «О»* — $\frac{2}{6}$.

Потом стало доподлинно известно, что произошло событие A_2. *Тогда вероятность вытащить «К»* — $\frac{1}{5}$.

Тогда $A = A_1 \cdot A_2 \cdot A_3 = \frac{1}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \dfrac{1}{105}$

2.1.8.3 Независимые события

Пусть A и B — два события, связанные с некоторым случайным экспериментом.

Определение. *События A и B называются независимыми, если* $P(AB) = P(A)\,P(B)$.

Теорема.

1. *Пусть* $P(B) > 0$.

   Утверждение «A и B — независимы» равносильно $P(A \,|\, B) = P(A)$;

2. *Пусть* $P(A) > 0$.

   Утверждение «A и B — независимы» равносильно $P(B \,|\, A) = P(B)$.

Доказательство.

1. *Сначала докажем, что если A и B — независимые, то* $P(A \,|\, B) = P(A)$. *По определению независимых событий,* $P(AB) = P(A)P(B)$. *По определению условной вероятности,*

   $$P(A \,|\, B) = \frac{P(AB)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

   Теперь докажем обратное.

   Пусть $P(A \,|\, B) = P(A)$. *Докажем, что* $P(AB) = P(A)P(B)$.

   $$P(AB) \overset{\text{по ф-ле умножения вероятностей}}{=} P(B) \cdot \overset{=P(A)}{P(A \,|\, B)} = P(B)P(A)$$

2. *Доказательство второго пункта теоремы аналогично.*

Замечание. *Разумеется, в качестве определения независимых событий логично было бы использовать условия*

$$P(A \mid B) = P(A) \text{ или } P(B \mid A) = P(B) \tag{6}$$

Однако эти условия имеют смысл лишь тогда, когда $P(A)$ или $P(B)$ отлично от нуля. Условие же $P(AB) = P(A)P(B)$ «работает» всегда без ограничений и, как мы показали выше, при выполнении соответствующих требований эквивалентно 6.

Пример. *Из колоды в 36 карт случайным образом извлекают карту.*

$$A = \{ \text{ извлечён туз } \}$$
$$B = \{ \text{ извлечена карта красной масти } \}$$

Являются ли A и B независимыми?

Условие независимости — $P(AB) = P(A)P(B)$.

Вероятности $P(A) = \frac{4}{36} = \frac{1}{9}$, $P(B) = \frac{18}{36} = \frac{1}{2}$; вероятность того, что был вытащен туз красной масти $P(AB) = \frac{2}{36} = \frac{1}{18}$.

Т. к. $P(A)\,P(B) = P(AB)$, то события A и B независимы.

Теорема. *Пусть события A, B независимы. Тогда независимыми являются события*

1. $\overline{A}$ *и* B;
2. A *и* $\overline{B}$;
3. $\overline{A}$ *и* $\overline{B}$.

Доказательство. *Проверим равенство*

$$P(\overline{A}B) = P(\overline{A})P(B) \tag{7}$$

1. *Если $P(B) = 0$, то правая часть 7 равна 0.*

$$\overline{A}B \subseteq B \implies P(\overset{\geqslant 0}{\overline{A}}B) \leqslant P(B) = 0 \implies P(\overline{A}B) = 0$$

Выяснив, что левая часть 7 равна 0, получается, что равенство верно.

2. *Если $P(B) > 0$:*

Тогда

$$P(\overline{A}B) \overset{\text{по теореме об умножении вероятностей}}{=} P(B) \cdot P(\overline{A} \mid B) =$$
$$\overset{\text{(т. к. условная вероятность обладает всеми свойствами безусловной вероятности)}}{=}$$
$$= P(B)(1 - P(A \mid B)) \overset{A,\ B \text{ независимы} \implies P(A|B)=P(A)}{=} P(B)(1 - P(A)) = P(B)P(\overline{A})$$

Остальное доказывается аналогично.

Определение. *События $A_1, \ldots, A_n$ называется попарно независимыми, если* [20]

$$\forall\forall\, i \neq j;\ i,\, j \in \{1, \ldots, n\}\ P\{A_i A_j\} = P\{A_i\} P\{A_j\}$$

Определение. *События $A_1, \ldots, A_n$ называются независимыми в совокупности, если*

$$\forall\, k \in \{2, \ldots, n\}\, \forall\forall\, i_1 < i_2 < \ldots < i_k\ P\{A_{i_1}, \ldots, A_{i_k}\} = P\{A_{i_1}\} \cdot \ldots \cdot P\{A_{i_k}\}$$

Замечание 1. *Условие из последнего определения означает, что должны выполняться следующие равенства*

$$\begin{aligned}
&k = 2\colon P\{A_{i_1}, A_{i_2}\} = P\{A_{i_1}\}\, P\{A_{i_2}\}\\
&k = 3\colon P\{A_{i_1}, A_{i_2}, A_{i_3}\} = P\{A_{i_1}\}\, P\{A_{i_2}\}\, P\{A_{i_3}\}\\
&\ldots\\
&k = n\colon \ldots
\end{aligned}$$

Замечание 2. *Очевидно, что если события $A_1, \ldots, A_n$ независимы в совокупности, то они и попарно независимы. Обратное неверно.*

Пример. ***(Бернштейна)***

Рассмотрим правильный тетраэдр[21]*, на одной грани которого «написано» 1, второй — 2, третьей — 3, четвёртой — 1, 2, 3.*

Этот тетраэдр один раз подбрасывают.

Событие A_1 заключается в том, что на нижней грани «написано» 1; также введём A_2 для 2, A_3 для 3. Давайте покажем, что события A_1, A_2, A_3 попарно независимы, но не являются независимыми в совокупности.

1. *Докажем, что они независимы попарно. Т. к. $P(A_1) = \frac{1}{2}$, $P(A_2) = \frac{1}{2}$, то*

   $$P(A_1 A_2) = P(A_1)\, P(A_2) = \frac{1}{4}$$

   Событие $A_1 A_2$ означает, что на нижней грани присутствуют и 1, и 2.

   Всё аналогично для $P(A_1 A_3) = P(A_1) P(A_3)$ и $P(A_2 A_3) = P(A_2) P(A_3)$.

2. *Проверим равенство $P(A_1 A_2 A_3) = P(A_1)\, P(A_2)\, P(A_3)$, которое, казалось бы, должно равняться $\frac{1}{8}$. Но произведение событий A_1, A_2, A_3 означает, что на нижней грани присутствуют и 1, и 2, и 3, вероятность чего равна $\frac{1}{4}$.*

   И выходит, что $\frac{1}{4} \neq \frac{1}{8}$.

   Следовательно, события A_1, A_2, A_3 не являются независимыми в совокупности.

---

[20] Обозначение $\forall\forall$ является математическим сленгом и технически некорректно. Тем не менее, это удобный способ обозначения того, что в выражении должно стоять несколько $\forall$ подряд. — Прим. лект.

[21] Трёхмерная фигура, состоящая из четырёх треугольников. — Прим. ред.

2.1.8.4 Формула полной вероятности

Пусть Ω — пространство элементарных исходов, связанных с некоторым случайным экспериментом, а $(\Omega,\, \beta,\, P)$ — вероятностное пространство этого случайного эксперимента.

Определение. *Говорят. что события $H_1,\, \ldots,\, H_n \in \beta$ образуют полную группу событий, если*

1. $P(H_i) > 0,\ i = \overline{1;\, n}$*;*

2. $H_i H_j = \emptyset$ *при* $i \neq j$*;*

3. $H_1 + \ldots + H_n = \Omega$*.*

Теорема. ***Формула полной вероятности.***
Пусть

1. $H_1,\, \ldots,\, H_n$ *— полная группа событий;*

2. $A \in \beta$ *— событие.*

Тогда (это выражение называется формулой полной вероятности):

$$P(A) = P(A \,|\, H_1) P(H_1) + \ldots + P(A \,|\, H_n) P(H_n)$$

Доказательство.

(см. рисунок 19)

1. $A = A\Omega \overset{\Omega = H_1 + \ldots + H_n}{=} A \cdot (H_1 + \ldots + H_n) = AH_1 + \ldots + AH_n$.
   Принимая $i \neq j$*:* $H_i \neq \emptyset$, $H_j \neq \emptyset$, *но* $(AH_i) \subseteq H_i$, $(AH_j) \subseteq H_j \implies (AH_i)(AH_j) = \emptyset$, *т. е.* AH_i *попарно не пересекаются.*

2. *Тогда*

$$P(A) = P(AH_1 + ... + AH_n) =$$
$$\overset{AH_i \text{ попарно не пересекаются}}{=}$$
$$= P(AH_1) + \ldots + P(AH_n) =$$
$$\overset{\text{т. к. } P(H_i) > 0, \text{ то } P(AH_i) = P(H_i) P(A \,|\, H_i)}{=}$$
$$= P(A \,|\, H_1) P(H_1) + \ldots + P(A \,|\, H_n) P(H_n)$$

Пример. *В мастерской продаются телевизоры трёх фирм. Первой произведено* 30%*, второй —* 50%*, третьей —* 20%*. Известно, что продукция первой фирмы содержит* 7% *процентов брака, второй —* 5%*, третьей —* 10%*. Какова вероятность того, что случайно выбранный телевизор окажется бракованным?*

Решение. *Событие A у нас будет заключаться в том, что случайно выбранный телевизор бракованный. Введём полную группу событий следующим образом:*

$$H_1 = \{ \text{ телевизор произведён первой фирмой } \}$$
$$H_2 = \{ \text{ телевизор произведён второй фирмой } \}$$
$$H_3 = \{ \text{ телевизор произведён третьей фирмой } \}$$

Можно воспользоваться формулой полной вероятности:

$$P(A) = P(A \,|\, H_1)P(H_1) + P(A \,|\, H_2)P(H_2) + P(A \,|\, H_3)P(H_3) = 0.066$$

Замечание. *События $H_1, \ldots, H_n$, образующие полную группу, часто называют гипотезами.*

2.1.8.5 Формула Байеса

Теорема. *Пусть*

1. $H_1, \ldots, H_n$ *— полная группа событий;*
2. $P(A) > 0$.

   Тогда

$$P(H_i \,|\, A) = \frac{P(A \,|\, H_i)P(H_i)}{P(A \,|\, H_1)P(H_1) + \ldots + P(A \,|\, H_n)P(H_n)},\ i = \overline{1;\, n}$$

Доказательство.

$$P(H_i \,|\, A) \overset{\text{по опр. условной вероятности}}{=}$$
$$= \frac{P(AH_i)}{P(A)} \overset{\text{по ф-ле умножения в числителе, полной вероятности в знаменателе}}{=}$$
$$= P(H_i \,|\, A) = \frac{P(A \,|\, H_i)P(H_i)}{P(A \,|\, H_1)P(H_1) + \ldots + P(A \,|\, H_n)P(H_n)},\ i = \overline{1;\, n}$$

Пример. *Пусть в условиях предыдущего примера о телевизорах известно, что куплен бракованный телевизор. Какова вероятность того, что он произведён второй фирмой? Какой фирмой он, скорее всего, произведён?*

$$P(H_1 \,|\, A) = \frac{P(A \,|\, H_1)P(H_1)}{P(A)} = \frac{0.021}{0.066} \approx 0.318$$
$$P(H_2 \,|\, A) = \frac{P(A \,|\, H_2)P(H_2)}{P(A)} = \frac{0.025}{0.066} \approx 0.379$$
$$P(H_3 \,|\, A) = \frac{P(A \,|\, H_3)P(H_3)}{P(A)} = \frac{0.02}{0.066} \approx 0.303$$

Ответ: вероятнее всего, бракованный телевизор произведён второй фирмой[22].

[22] Обратите внимание, что у второй фирмы наименьший процент брака. Результат связан с тем, что она в то же время производит наибольшее кол-во телевизоров. — Прим. лект.

Вероятности $P(H_i)$, $i = \overline{1;\, n}$ называются априорными, т. к. они известны до опыта; Вероятности $P(H_i \mid A)$, $i = \overline{1;\, n}$ называются апостериорными — они вычисляются после опыта.

Лекция №7, 16.10.2018

2.1.8.6 Схемы испытаний Бернулли

Давайте рассмотрим случайный эксперимент, в результате которого возможна реализация одного из двух элементарных исходов, т. е. пространство элементарных исходов у нас будет состоять из двух элементов ($|\Omega| = 2$).

Один из элементарных исходов условно будем называть успехом, второй — неудачей. Пусть p — вероятность осуществления успеха в случайном эксперименте, а q ($q = 1 - p$) — вероятность неудачи.

Определение. *Схемой испытаний Бернулли называется серия из однотипных экспериментов указанного вида, в которой отдельные испытания независимы, т. е. вероятность реализации успеха в i-ом испытании не зависит от исходов первого, второго, ..., $i - 1$-ого испытаний.*

Пример. *n раз подбрасывают игральную кость. Успехом будем считать выпадение 6-ки, а неудачей — всё остальное. Тогда $p = \frac{1}{6}$, $q = \frac{5}{6}$.*

Пример. *В роддоме наблюдают очередного новорождённого ребёнка. Будем считать, что «успех» — это рождение мальчика, а «неуспех» — рождение девочки. Тогда $p \approx q \approx \frac{1}{2}$.*

Пример. *Покупают n лотерейных билетов некоторого тиража и выбирают последовательно билеты; «успех» — вытаскивание выигрышного билета. Условия не удовлетворяют схеме Бернулли, т. к. испытания не являются независимыми. Например, предполагая, что куплено $n = 100$ билетов, то с каждым вытащенным билетом вероятности будут изменяться. Тем не менее, если n — кол-во купленных билетов и оно много меньше N — кол-ва билетов в общем тираже ($n \ll N$), то схема испытаний Бернулли удовлетворительно описывает рассматриваемый эксперимент.*

Теорема. *Пусть проводится серия из n испытаний по схеме Бернулли с вероятностью успеха p. Тогда $P_n(k)$ есть вероятность того, что в серии из n испытаний произойдёт ровно k успехов:*

$$P_n(k) = C_n^k p^k q^{n-k}$$

Доказательство.

1. *Результат проведения серии из n экспериментов запишем с использованием кортежа $(x_1, \ldots, x_n)$, где*

$$x_i = \begin{cases} 1, \text{ если в испытании имел место успех;} \\ 0, \text{ если в испытании имела место неудача.} \end{cases}$$

2. *Пусть*

$$A = \{ \text{ в серии из } n \text{ испытаний произошло ровно } k \text{ успехов } \}$$

Тогда A состоит из кортежей, в которых будет ровно k единиц и $n-k$ нулей.

В событии A будет столько элементарных исходов, сколькими способами можно расставить k единиц по n позициям. Каждая такая расстановка однозначно определяется номерами позиций, в которых будут записаны единички. В остальные позиции будут записаны нули.

Выбрать k позиций из имеющихся n можно C_n^k способами. Вероятность каждого отдельного исхода равна произведению вероятностей каждого отдельного x_i, и тогда общая вероятность исхода будет равна $p^k q^{n-k}$.

Все испытания независимы; следовательно, все кортежи из A равновероятны, и их C_n^k штук, что означает

$$P_n(k) = C_n^k p^k q^{n-k}$$

Следствие. *Вероятность того, что кол-во успехов в серии из n испытаний по схеме Бернулли с вероятностью успеха p будет заключено между k_1 и k_2:*

$$P_n(k_1 \leqslant k \leqslant k_2) = \sum_{i=k_1}^{k_2} C_n^i p^i q^{n-i}$$

Доказательство.

1. *Пусть $A_i = \{ \text{ в серии произошло ровно } i \text{ успехов } \}$, $i = \overline{k_1;\, k_2}$.*
   $P(A_i) = P_n(i) = C_n^i p^i q^{n-i}$.

2. *Вероятность $P_n(k_1 \leqslant k \leqslant k_2)$ соответствует событию $A = A_{k_1} + A_{k_1+1} + \ldots + A_{k_2}$, выразим*

$$P(A) = P(A_{k_1} + \ldots + A_{k_2}) \overset{A_i \text{ и } A_j \text{ несовместны при } i \neq j}{=} P(A_{k_1}) + \ldots + P(A_{k_2}) =$$
$$= \sum_{i=k_1}^{k_2} P\{A_i\} = \sum_{i=k_1}^{k_2} C_n^i p^i q^{n-i}$$

Следствие. *Вероятность того, что в серии испытаний Бернулли с вероятностью успеха p (и неудачи $q = 1-p$) произойдёт хотя бы один успех: $P_n(k \geqslant 1) = 1 - q^n$.*

Доказательство. *Пусть $A = \{ \text{ в серии произошёл хотя бы один успех } \}$. В таком случае $\overline{A} = \{ \text{ в серии не будет ни одного успеха } \}$, и тогда*

$$P(A) = 1 - P(\overline{A}) = 1 - P_n(0) = 1 - C_n^0 p^0 q^{n-0} = 1 - q^n$$

Пример. *Игральную кость бросают пять раз. Найти вероятности следующих событий:*

$$A = \{ \text{ шестёрка выпадет ровно два раза } \}$$
$$B = \{ \text{ шестёрка выпадет хотя бы два раза } \}$$

Задав «удача» как выпадение шестёрки, получим

1. $P(A) = P_5(2) \approx 0.161$.

2. $$P(B) = P(k \geqslant 2) = \sum_{i=2}^{5} C_5^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2} \quad (8)$$

Можно заметить, что вычисление выражения 8 сложно. В таком случае можно попытаться высчитать отрицание

$$P(\overline{B}) = P_5(0 \leqslant k \leqslant 1) = C_5^0(\tfrac{1}{6})^0(\tfrac{5}{6})^5 + C_5^1(\tfrac{1}{6})^1(\tfrac{5}{6})^4$$

и тогда $P(B) = 1 - P(\overline{B}) \approx 0.196$.

2.2 Случайные величины

2.2.1 Одномерные случайные величины

2.2.1.1 Понятие случайной величины

Нестрогое определение. *Пусть исход случайного эксперимента можно описать числом X. Тогда X — случайная величина.*

Пример. *Рассмотрим величину X в различных ситуациях:*

1. *Один раз подбрасывают игральную кость, X — кол-во очков; тогда $X \in \{1, \ldots 6\}$ — случайная величина;*
2. *Проводят серию из n экспериментов по схеме испытаний Бернулли, X — кол-во успехов в этой серии; $X \in \{0, 1, \ldots, n\}$;*
3. *Подбрасывают симметричную монету до первого появления герба, X — кол-во подбрасываний; тогда $X \in \{1, 2, 3, \ldots\} = N$;*
4. *У случайного выбранного пациента больницы измеряют температуру тела X. Тогда $X \in [34; 41]$;*
5. *Производят стрельбу по плоской мишени, X — расстояние от центра мишени до точки попадания пули; тогда $X \in [0; +\infty]$.*

Пусть (Ω, β, P) — вероятностное пространство некоторого случайного эксперимента.

Определение. *Случайной величиной называется функция*

$$X \colon \Omega \to \mathbb{R}$$

такая, что для каждого $x \in \mathbb{R}$ множество $\{\omega : X(\omega) < x\} \in \beta$ (т. е. для любого x множество $\{\omega \colon X(\omega) < x\}$ является событием).

Замечание. *Упрощённо на случайную величину можно смотреть как на случайный эксперимент, в котором на прямую бросают точку. Точка x называется реализацией величины X.*

Пример. *Обозначим величину X, используя ситуации ранее приведённых примеров.*

1. *Если подбрасывают игральную кость, то будут выбираться точки 1, 2, 3, 4, 5, 6 (каждая с вероятностью $\frac{1}{6}$).*

2. *(пример был проигнорирован на лекции)*

3. *Точка 1 будет выбираться с вероятностью* $\frac{1}{2}$, *2* — $\frac{1}{4}$, *3* — $\frac{1}{8}$, *4* — $\frac{1}{16}$ *и т. д.*

4. *См. рисунок 20.*

5. *См. рисунок 21.*

Во всех разобранных выше примерах случайные величины имели различные диаграммы распределения частот своих значений.

Определение. *Правило, в соответствии с которым различные возможные значения (множества значений) случайной величины приписываются вероятности того, что случайная величина примет эти значения (примет значения из множества), называются законом распределения случайной величины.*

Универсальным способом задания закона распределения случайной величины является функция распределения.

2.2.1.2 Функция распределения

Пусть X — случайная величина, связанная с некоторым случайным экспериментом.

Определение. *Функцией распределения вероятностей случайной величины X называется отображение*

$$F_X \colon \mathbb{R} \to \mathbb{R}$$

определённое следующим правилом[23]

$$F_X(x) = P\{X < x\}$$

Замечание 1. *Из определения функции распределения следует, что* $\forall x \in \mathbb{R}$ *множество* $\{\omega : X(\omega) < x\}$ *должно быть событием (ведь вероятность P определена только для элементов из β). Но это условие выполнено в силу определения случайной величины.*

Замечание 2. *Значение функции распределения F_X в точке x равно вероятности того, что случайно брошенная на прямую точка попадёт левее x.*

Пример. *Два раза бросают симметричную монету. Пусть X — кол-во выпадения герба при этих подбрасываниях. Просят построить функцию распределения случайной величины X.*

Возможные значения X: $X \in \{0, 1, 2\}$

[23] С тем же успехом можно определить F_X как $F_X(x) = P\{X \leqslant x\}$; выбор обусловлен конвенцией (при этом, видимо, в международной литературе используют именно знак $\leqslant$). От того, как определена эта функция, в основном зависят выборы знаков ($<$ или $\leqslant$) в определённых местах в дальнейших формулах, но список отличий этим не ограничивается. — Прим. ред.

$$\begin{array}{cccc} X & 0 & 1 & 2 \\ P & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}$$

(график функции смотри на рисунке 22)

$$F_X(x) = \begin{cases} 0, & x \in (-\infty;\, 0] \\ \frac{1}{4}, & x \in (0;\, 1] \\ \frac{3}{4}, & x \in (1;\, 2] \\ 1, & x \in (2; +\infty) \end{cases}$$

Обратите особое внимание на интервалы (а точнее, на их границы).

Лекция №8, 23.10.2018

2.2.1.2.1 Свойства функции распределения

1. $0 \leqslant F(x) \leqslant 1$;
2. F является неубывающей функцией, т. е. если $x_1 \leqslant x_2$, то $F(x_1) \leqslant F(x_2)$;
3. $$\lim_{x\to-\infty} F(x) = 0$$ $$\lim_{x\to+\infty} F(x) = 1$$
4. В каждой точке функция распределения непрерывна слева[24]: $\lim\limits_{x\to x_0-} F(x) = F(x_0)$;
5. $P\{a \leqslant X < b\} = F(b) - F(a)$.

Доказательства.

1. *$F(x)$ определена как вероятность, т. е. $F(x) = P\{...\} \in [0;\, 1]$.*
2. *Имея $x_1 \leqslant x_2$, выразим $F(x_2)$:*

$$F(x_2) = P\{X < x_2\} = P\{\underbrace{\{X < x_1\}}_{\text{Событие } A} + \underbrace{\{x_1 \leqslant X < x_2\}}_{\text{Событие } B}\}$$

События A и B несовместны, т. е.

$$F(x_2) = \underbrace{P\{X < x_1\}}_{F(x_1)} + \underbrace{P\{x_1 \leqslant X < x_2\}}_{\geqslant 0} \geqslant F(x_1)$$

[24] Если бы F_X была определена как $F_X = P\{X \leqslant x\}$, то функция распределения была бы непрерывна справа. В учебнике по Теории Вероятностей из серии «Математика в техническом университете» (с римской цифрой на обложке, далее книги из серии будут адресоваться по номерам; для Теории Вероятностей — учебник XVI) написано (издание третье, исправленное), что на этом отличия в свойствах заканчиваются; это, судя по всему, ошибка, т. к. при использовании $\leqslant$ свойство 5 должно записываться как $P\{a < X \leqslant b\} = F(b) - F(a)$. — Прим. ред.

3. *Сначала докажем, что* $\lim\limits_{x\to+\infty} F(x) = 1$*:*

   Рассмотрим последовательность $x_1, x_2, x_3, \ldots$ *такую, что*

   (a) $x_1 < x_2 < x_3 < \ldots$;

   (b) $\lim\limits_{n\to\infty} x_n = +\infty$.

   Обозначим $A_n = \{X < x_n\}$, $n \in N$*; очевидно, что последовательность событий* A_n *имеет свойство* $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq A_{n+1} \subseteq \ldots$, *т. е. эта последовательность является неубывающей последовательностью событий.*

   Тогда, применяя аксиому непрерывности

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(A_n) \overset{\text{аксиома непрерывности}}{=} \underbrace{P\{X < +\infty\}}_{\text{(достоверное событие)}} = 1$$

   Т. к. $x_1, x_2, \ldots$ *— произвольная последовательность (неубывающая и стремящаяся к бесконечности), то в соответствии с определением предела функции по Гейне*[25]

$$\lim_{x\to+\infty} F(x) = 1$$

   Другая часть этого свойства, $\lim\limits_{x\to-\infty} F(x) = 0$, *доказывается аналогично.*

4. *Пусть* $x_1, x_2, \ldots$ *— возрастающая последовательность такая, что* $\lim\limits_{n\to\infty} x_n = x_0$.

   Пусть $A_i = \{X < x_i\}$, $i \in \mathbb{N}$. *Тогда событие*

$$\{X < x_n\} = \bigcup_{i=1}^{\infty} A_i$$

   причём последовательность событий $A_1, A_2, \ldots$ *является возрастающей;*

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(A_n) \overset{\text{аксиома непрерывности}}{=} P\{X < x_0\} = F(x_0)$$

   Т. к. $x_1, x_2, \ldots$ *— произвольная последовательность, сходящаяся к* x_0 *слева, то в соответствии с определением предела функции по Гейне*

$$\lim_{x\to x_0-} F(x) = F(x_0)$$

[25] Значение A называется пределом функции $f(x)$ в точке x_0, если для любой последовательности точек $\{x_n\}_{n=1}^{\infty}$, сходящейся к x_0 (см. понятие предела *последовательности*, не функции), но не содержащей x_0 в качестве одного из своих элементов, последовательность значений функции $\{f(x_n)\}_{n=1}^{\infty}$ сходится к A. — Прим. ред.

5. $\{X < b\} = \{X < a\} + \{a \leqslant X < b\}$; *события в объединении несовместные, поэтому*

$$\underbrace{P\{X < b\}}_{F(B)} = \underbrace{P\{X < a\}}_{F(A)} + P\{a \leqslant X < b\}$$

из чего тривиально следует

$$P\{a \leqslant X < b\} = F(b) - F(a)$$

Замечание. *Можно показать, что любая функция, обладающая свойствами 2, 3 и 4 является функцией распределения некоторой случайной величины.*

2.2.1.3 Дискретные случайные величины

Определение. *Случайная величина называется дискретной, если множество её значений конечно или счётно.*

Пусть X — дискретная случайная величина, принимающая значения из конечного множества $\{x_1, \ldots, x_n\}$. Закон распределения такой случайной величины можно задать таблицей $P(x)$:

X	x_1	$\ldots$	x_i	$\ldots$	x_n
P	p_1	$\ldots$	p_i	$\ldots$	p_n

При этом $p_i = P\{X = x_i\},\ i = \overline{1;\,n};\ \sum\limits_{i=1}^{n} p_i = 1.$

Определение. *Эта таблица называется рядом распределения дискретной случайной величины X.*

Пример. ***Задача.*** *Пусть X — кол-во бросков монеты до первого появления герба. Составить ряд распределения случайной величины X.*

Решение. *$X \in N$ — счётное множество, т. е. X — дискретная величина.*

X	1	2	$\ldots$	n	$\ldots$
P	$\frac{1}{2}$	$\frac{1}{4}$	$\ldots$	$\frac{1}{2^n}$	$\ldots$

$P\{X = k\} = \dfrac{1}{2^k},\ k \in \mathbb{N}.$

Проверка.

$$\sum_{k=1}^{\infty} P\{X = k\} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = \frac{1}{2}\left(1 + \frac{1}{2} + \frac{1}{4}\right) =$$

$$\overset{\text{по формуле суммы бесконечно убывающей геометрической прогрессии } S=\frac{b_1}{1-q} \text{ при } q=\frac{1}{2},\, b_1=1}{=}$$

$$= \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2}} = 1$$

2.2.1.4 Непрерывные случайные величины

Определение. *Случайная величина X называется непрерывной, если существует функция*

$$f(x)\colon \mathbb{R} \to \mathbb{R}$$

такая, что $\forall x \in \mathbb{R}$ функция $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ (F — функция распределения в X).

(см. рисунок 23)

Замечание 1. *При этом f называют функцией плотности распределения вероятности случайной величины X.*

Замечание 2. *Для большинства представляющих практический интерес непрерывных случайных величин функция плотности f является непрерывной или кусочно-непрерывной. Это означает, что функция распределения $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ является непрерывной функцией (это обстоятельство и объясняет термин «непрерывная» случайная величина).*

Замечание 3. *Если известна плотность f, то понятно, как найти функцию распределения:*

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt \tag{9}$$

Обратно, в соответствии с теоремой о производной интеграла с переменным верхним пределом[26]*,*

$$f(x) = F'(x) \tag{10}$$

для всех точек $x \in \mathbb{R}$, в которых f непрерывна (т. е. почти для всех x).

Из 9 и 10 следует, что если известна одна из функций F или f, то можно найти и другую. Плотность распределения случайной величины также содержит всю информацию о законе распределения этой случайной величины. Таким образом, закон распределения можно задавать как с использованием функции распределения F, так и с использованием функции плотности f.

2.2.1.4.1 Свойства непрерывных случайных величин

1. $f(x) \geqslant 0$;

2. $$P\{a \leqslant X < b\} = \int\limits_{a}^{b} f(x)dx$$

---

[26] См. учебник VI «Интегральное исчисление функций одного переменного»; описанное не оформлено как теорема, но упоминается на стр. 247. — Прим. ред.

где X — непрерывная случайная величина, а f – её функция плотности[27];

3. Условие нормировки:
$$\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1$$
где f — функция плотности некоторой случайной величины;

4.
$$P\{x_0 \leqslant X < x_0 + \triangle x\} \approx f(x_0)\triangle x$$
где X — непрерывная случайная величина, f — её функция плотности, x_0 — точка непрерывности функции f, а $\triangle x$ — мало;

5. Если X — непрерывная случайная величина, то для любого наперёд заданного $x_0 \in \mathbb{R}$
$$P\{X = x_0\} = 0$$

Доказательства.

1. *Почти всюду* $f(x) = F'(x) \overset{F\ -\ \textit{неубывающая}}{\geqslant} 0.$

2. *По свойству функции распределения*
$$P\{x \leqslant X < b\} = F(b) - F(a)$$
Т. к. F *— первообразная для* f*, то по формуле Ньютона-Лейбница*
$$P\{x \leqslant X < b\} = F(b) - F(a) = \int\limits_a^b f(x)dx$$

3.
$$\int\limits_{-\infty}^{+\infty} f(x)\,dx = \lim_{x_1\to-\infty,\, x_2\to+\infty} \int\limits_{x_1}^{x_2} f(x)\,dx = \lim_{x_2\to+\infty} F(x_2) - \lim_{x_1\to-\infty} F(x_1) =$$
$$= \cancelto{1}{F(+\infty)} - \cancelto{0}{F(-\infty)} = 1$$

[27]Чуть позже по лекциям утверждается, что в этом свойстве не важно, какие знаки стоят в условии при P, т. е. $P\{a \leqslant X < b\} = P\{a < X < b\} = P\{a < X \leqslant b\} = P\{a \leqslant X \leqslant b\} = P\{a \preccurlyeq X \preccurlyeq b\} = \int\limits_a^b f(x)dx$. — Прим. ред.

4.

$$P\{x_0 \leqslant X < x_0 + \triangle x\} \overset{\text{свойство 5}}{=} F(x_0 + \triangle x) - F(x_0)$$

Т. к. x_0 — точка непрерывности f, а $\triangle x$ мало, то можно считать, что в окрестности $(x_0, x_0 + \triangle x)$ функция $F' = f$ непрерывна. Тогда применим к функции f на $[x_0, x_0 + \triangle x]$ теорему Лагранжа[28] [29]*:*

$$F(x_0 + \triangle x) - F(x_0) = \underbrace{F'(\xi)}_{f(\xi)} \triangle x$$

где $\xi \in (x_0, x_0 + \triangle x)$. Т. к. $\triangle x$ мало, а f непрерывна в некоторой окрестности x_0, то можно считать, что $f(\xi) \approx f(x_0)$. Таким образом,

$$P\{x_0 \leqslant X < x_0 + \triangle x\} \approx f(x_0)\triangle x$$

5.

$$P\{X = x_0\} = \lim_{\triangle x \to 0} P\{x_0 \leqslant X < x_0 + \triangle x\} =$$

$$\overset{\text{свойство 2}}{=}$$

$$= \lim_{\triangle x \to 0} [F(x_0 + \triangle x) - F(x_0)] \overset{F \text{ непрерывна, см. замечание выше}}{=} 0$$

Замечание 1. *В силу свойства 5 свойство 2 можно записать в следующем виде:*

Если X — непрерывная величина, то

$$P\{a \leqslant X < b\} = P\{a < X < b\} = P\{a < X \leqslant b\} = P\{a \leqslant X \leqslant b\} = \int\limits_a^b f(x)dx$$

Условие события такого вида в дальнейшем будем записывать в виде

$$\{a \preccurlyeq X \preccurlyeq b\}$$

Замечание 2. *Дискретные и непрерывные случайные величины являются в некотором смысле «крайними» моделями случайных величин. На практике встречаются такие случайные величины, которые называются комбинированными или смешанного типа.*

Лекция №9, 30.10.2018

[28]Пусть функция $y = f(x)$ непрерывна на отрезке $[a, b]$ и дифференцируема в интервале (a, b); тогда между точками a и b найдётся хотя бы одна такая точка c $(a < c < b)$, для которой справедливо равенство $f(b) - f(a) = f'(c)(b - a)$. — Прим. ред.

[29]ξ — строчная буква «кси» греческого алфавита. — Прим. ред.

2.2.1.5 Основные законы распределения случайных величин

2.2.1.5.1 Пуассоновская случайная величина

Определение. *Говорят, что случайная величина X распределена по закону Пуассона с параметром $\lambda > 0$, если она принимает значения $0,\ 1,\ 2, \ldots$ с вероятностями*

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda},\ k \in \{0,\ 1,\ 2,\ \ldots\}$$

Обозначается $X \sim \Pi(\lambda)$.

Замечание 1. *Ряд распределения случайной величины X выглядит как*

$$\begin{array}{cccccc} k & 0 & 1 & \ldots & k & \ldots \\ P\{X = k\} & e^{-\lambda} & \lambda e^{-\lambda} & \ldots & \dfrac{\lambda^k}{k!} e^{-\lambda} & \ldots \end{array}$$

Проверим условие нормировки:

$$\sum_{k=0}^{\infty} P\{X = k\} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\text{ряд Маклорена для } e^{\lambda}} = e^{\lambda} e^{-\lambda} = 1$$

Замечание 2. *Распределение Пуассона называют законом редких событий, т. к. оно проявляется там, где происходит большое число испытаний с малой вероятностью успеха. Например, кол-во метеоритов, упавших в данной местности за данный промежуток времени распределено по закону Пуассона (или близкому к нему закону).*

2.2.1.5.2 Биномиальная случайная величина

Определение. *Говорят, что случайная величина X распределена по биномиальному закону с параметрами $n \in N$ и $p \in (0; 1)$, если она принимает значения $0,\ 1,\ \ldots,\ n$ с вероятностями*

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k},\ k \in \{0,\ 1,\ \ldots,\ n\}$$

Обозначается $X \sim B(n,\ p)$.

Замечание 1. *Ряд распределения случайной величины X будет выглядеть как (здесь $q = 1 - p$):*

$$\begin{array}{cccccc} k & 0 & \ldots & k & \ldots & n \\ P\{X = k\} & q^n & \ldots & C_n^k p^k q^{n-k} & \ldots & p^n \end{array}$$

Проверим условие нормировки:

$$\sum_{k=0}^{\infty} P\{x=k\} = \underbrace{\sum_{k=0}^{\infty} C_n^k p^k q^{n-k}}_{\text{формула разложения бинома Ньютона}^{30}} = (p+q)^n = 1^n = 1$$

Замечание 2. *Случайная величина $X \sim B(n, p)$ — кол-во успехов в серии из n испытаний по схеме Бернулли с вероятностью p успеха в одном испытании.*

2.2.1.5.3 Геометрическое распределение

Определение. *Говорят, что случайная величина X имеет геометрическое распределение с параметром $p \in (0, 1)$, если она принимает значения $0, 1, \ldots$ с вероятностями (здесь $q = 1 - p$)*

$$P\{X = k\} = pq^k,\ k = 0, 1, 2, 3, \ldots$$

Обозначается[31] *как $X \sim Geom(p)$.*

Замечание 1. *Ряд распределения такой случайной величины имеет следующий вид:*

k	0	1	...	k	...
$P\{X=k\}$	p	pq	...	pq^k	...

Проверим условие нормировки:

$$\sum_{k=0}^{\infty} P\{x=k\} = \sum_{k=0}^{\infty} pq^k = p \underbrace{\sum_{k=0}^{\infty} q^k}_{\text{бесконечно убывающая геометрическая прогрессия}} = p \cdot \frac{1}{\underbrace{1-q}_{=p}} = 1$$

Замечание 2. *Случайная величина X, имеющая геометрическое распределение с параметром p — это кол-во испытаний по схеме Бернулли (с вероятностью p успеха в одном испытании), которое нужно произвести **до** первого появления успеха (т. е. если в серии успех впервые произошёл в k-ом испытании, то $X = k - 1$).*

[30] Формула бино́ма Ньютона (англ. Binomial theorem, Binomial expansion) — формула для разложения на отдельные слагаемые целой неотрицательной суммы двух переменных в степени n: $(a+b)^n = C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + \ldots + C_n^n a^0 b^n$. — Прим. ред.

[31] Обозначение геометрического распределения не было дано на лекциях. Указанное взято из соответствующей статьи на Википедии (статьи на русском — в английской используется Pr). — Прим. ред.

В самом деле, вероятность потерпеть неудачу k раз подряд и после этого получить успех равна (т. к. испытания независимые, то и события независимые)

$$P\{\underbrace{(n,n,n,n,)}_{k\ \text{раз}}y\} = P\{\text{ в первом испытании — неудача }\}\cdot$$
$$\cdot P\{\text{ во втором — неудача }\}\cdot$$
$$\cdot\ldots\cdot$$
$$\cdot P\{\text{ в } k\text{-ом — неудача }\}\cdot$$
$$\cdot P\{\text{ в } (k+1)\text{-ом успех }\}$$

что означает

$$P\{\underbrace{(n,n,n,n,)}_{k\ \text{раз}}y\} = \underbrace{q\cdot q\cdot\ldots\cdot q}_{k\ \text{раз}}\cdot p$$

Замечание. *Пуассоновская, биномиальная и геометрическая случайные величины являются дискретными случайными величинами.*

2.2.1.5.4 Равномерно распределённая случайная величина

Определение. *Говорят, что случайная величина X имеет равномерное распределение на отрезке $[a;\,b]$, если её функция плотности имеет вид*

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

Значение константы с однозначно определяется из условия нормировки.

$$1 = \int\limits_{-\infty}^{+\infty} f(x)\,dx = \int\limits_a^b c\,dx = c(b-a) = 1 \implies c = \frac{1}{b-a}$$

Обозначается $X\sim R[a;\,b]$.

Проверим условие нормировки:

(см. рисунок 24)

Замечание 1. *Пусть $X\sim R[a;b]$ и заданы α, β такие, что $a\leqslant\alpha<\beta\leqslant b$.*

По свойству непрерывной случайной величины,

$$P\{X\in[\alpha;\beta]\} = \int\limits_\alpha^\beta f(x)\,dx \overset{[\alpha;\beta]\subseteq[a;b]}{=} \int\limits_\alpha^\beta \frac{dx}{b-a} = \frac{\beta-\alpha}{b-a}$$

Замечание 2. *Таким образом, вероятность того, что равномерно распределённая величина примет значение из некоторого множества M пропорциональна мере этого множества ($M \subseteq [a; b]$). По этой причине равномерно распределённая величина реализует геометрическое определение вероятности в одномерном случае.*

2.2.1.5.5 Экспоненциальная случайная величина

Определение. *Говорят, что случайная величина X распределена по экспоненциальному закону с параметром $\lambda > 0$, если её плотность распределения имеет вид*

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

Обозначается $X \sim Exp(\lambda)$.

Замечание 1. *Условие нормировки:*

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

(см. рисунок 25)

$$F(x) = \int\limits_{-\infty}^{x} f(t)\,dt = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

Замечание 2. *Для многих технических устройств время X их безотказной работы распределено по экспоненциальному закону (при некотором подходящем значении параметра λ).*

2.2.1.5.6 Нормальная случайная величина

Определение. *Говорят, что случайная величина X имеет нормальное распределение с параметрами m и σ^2 ($\sigma > 0$), если её функция плотности имеет вид*

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

Обозначается $X \sim N(m, \sigma^2)$.

Функция плотности нормального распределения имеет характерную колоколообразную форму; m является координатой x «центра» этого колокола (центра симметрии), а σ характеризует разброс значений случайной величины; чем меньше σ, тем выше экстремум функции плотности[32].

Замечание 1. *График функции нормировки случайной величины:*

(см. рисунок 26)

Замечание 2. *Проверим условие нормировки; интеграл от $f(x)$ для нормального распределения в общем случае не берётся, но нам для нашей задачи нужно найти значение конкретного определённого несобственного интеграла, что можно сделать, применив трюк:*

1.

$$\int\limits_{-\infty}^{+\infty} f(x)\,dx = \int\limits_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-m)^2}{2\sigma^2}}\,dx =$$

$$= \left\langle t = \frac{x-m}{\sigma\sqrt{2}} \Leftrightarrow x = t\sigma\sqrt{2} + m;\; dx = \sigma\sqrt{2}\,dt, \begin{cases} x = -\infty \implies t = -\infty \\ x = +\infty \implies t = +\infty \end{cases} \right\rangle =$$

$$= \frac{1}{\sqrt{\pi}} \underbrace{\int\limits_{-\infty}^{+\infty} e^{-t^2}\,dt}_{\text{неберущийся интеграл, обозначим его как } I_0}$$

2. *Рассмотрим интеграл I*

$$I = \iint\limits_{R^2} e^{-x^2-y^2}\,dx\,dy = \int\limits_{-\infty}^{+\infty} e^{-x^2}\,dx \int\limits_{-\infty}^{+\infty} e^{-y^2}\,dy =$$

$$= \left(\int\limits_{-\infty}^{+\infty} e^{-x^2}\,dx\right)\left(\int\limits_{-\infty}^{+\infty} e^{-y^2}\,dy\right) = I_0^2$$

[32] В статье «Нормальное распределение» в Википедии присутствуют неплохие иллюстрации. — Прим. ред.

3.

$$I = \langle \text{ перейдём в полярную систему координат } \rangle =$$

$$= \iint\limits_{D_{\rho\varphi}} e^{-\rho^2}\rho\,d\rho\,d\varphi = \underbrace{\int\limits_0^{2\pi} d\varphi}_{=2\pi} \int\limits_0^{+\infty} e^{-\rho^2}\rho\,d\rho = \frac{2\pi}{2}\int\limits_0^{+\infty} e^{-\rho^2}\,d\rho^2 =$$

$$= \pi\left(-e^{-\rho^2}\right)\Big|_0^{+\infty} = 0 - (-\pi) = \pi$$

4. $I = \pi = I_0^2 \implies I_0 = \sqrt{\pi}$, *и тогда*

$$\int\limits_{-\infty}^{+\infty} f(x)\,dx = \int\limits_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-m)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{\pi}} \underbrace{\int\limits_{-\infty}^{+\infty} e^{-t^2}\,dt}_{I_0=\sqrt{\pi}} = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1$$

Замечание 3. *Функция распределения нормальной случайной величины* $X \sim N(m,\,\sigma^2)$*:*

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\dfrac{(t-m)^2}{2\sigma^2}}\,dt$$

Интеграл в формуле не берётся; это означает, что функция $F(x)$ *не является элементарной (т. е. не может быть задана одной формулой с использованием основных тригонометрических функций и операций сложения, умножения, деления и композиции над ними).*

Замечание 4.

Определение. *Распределение* $N(0,\,1)$ *называют стандартным нормальным распределением; для него функция плотности равна*

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}},\ x \in \mathbb{R}$$

Часто для вычисления вероятностей из стандартного нормального распределения рассматривают функцию

$$\Phi(x) = F_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$$

$\Phi(x)$ называют функцией Лапласа; для нахождения её значений используйте заранее высчитанные таблицы[33].

(см. рисунок 27)

Замечание 5. *Часто вместо функции $\Phi(x)$ удобнее рассмотреть функцию*

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$$

(см. рисунок 28)

Свойства функций Φ и Φ_0:

1. $\Phi(x) = \frac{1}{2} + \Phi_0(x)$;
2. $\Phi_0(-x) = -\Phi_0(x)$ *(функция нечётная);*
3. $\Phi_0(+\infty) = \frac{1}{2}$;
4. $\Phi_0(-\infty) = -\frac{1}{2}$.

Доказательства этих свойств:

1.

$$\Phi(x) = \int\limits_{-\infty}^{x} f_{0,1}(t)\,dt = \cancelto{\frac{1}{2}}{\int\limits_{-\infty}^{0} f_{0,1}(t)\,dt} + \underbrace{\int\limits_{0}^{x} f_{0,1}(t)\,dt}_{\Phi_0(x)} = \frac{1}{2} + \Phi_0(x)$$

[33]В конце учебника XVI (приложение П.3) находится таблица значений для функции $\Phi_0(x)$ (см. следующее замечание). Лектор говорит, что в домашнем задании в обязательном порядке необходимо посчитать вероятности при помощи этой функции; на контрольную или экзамен можно принести распечатку таблицы значений, но там высчитывание конечного значения вероятности не обязательно. — Прим. ред.

2.

$$\Phi_0(-x) = \int\limits_0^{-x} f_{0,1}(t)\,dt =$$

$$= \left\langle t = -y,\ dt = -dy;\ \begin{cases} t = 0 \implies y = 0, \\ t = -x \implies y = x \end{cases} \right\rangle =$$

$$= -\int\limits_0^{x} f_{0,1}(-y)\,dy = \left\langle f_{0,1}(-y) = f_{0,1}(y) \right\rangle = -\int\limits_0^{x} f_{0,1}(y)\,dy = -\Phi_0(x)$$

3. По свойству 1 имеем $\Phi_0(x) \equiv \Phi(x) - \frac{1}{2}$*, из этого:*

$$\lim_{x\to+\infty} \Phi_0(x) = \underbrace{\lim_{x\to+\infty} \Phi(x) \to 1}_{\text{по свойству функции распределения}} - \frac{1}{2} = 1 - \frac{1}{2} = \frac{1}{2}$$

4. Аналогично предыдущему доказательству:

$$\lim_{x\to-\infty} \Phi_0(x) = \underbrace{\lim_{x\to-\infty} \Phi(x) \to 0}_{\text{по свойству функции распределения}} - \frac{1}{2} = 0 - \frac{1}{2} = -\frac{1}{2}$$

Лекция №10, 06.11.2018

Замечание 6. *Пусть* $x \sim N(m,\, \sigma^2)$*; чему равно* $P\{a \preccurlyeq X \preccurlyeq b\}$*?*

Рассмотрим

$$P\{a \leqslant X < b\} = \langle \text{ по свойству плотности распределения } \rangle =$$

$$= \int\limits_a^b f_X(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_a^b e^{-\frac{(x-m)^2}{2\sigma^2}} =$$

$$= \left\langle t = \frac{x-m}{\sigma};\; dx = \sigma\,dt;\; \begin{cases} x = a \implies t = \frac{a-m}{\sigma}, \\ x = b \implies t = \frac{b-m}{\sigma} \end{cases} \right\rangle =$$

$$= \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \cdot \cancel{\sigma} \cdot \int\limits_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} e^{-\frac{t^2}{2}}\,dt = \langle \text{ т. к. } \Phi(t) \text{ — первообразная } f_{0,1}(t) \rangle =$$

$$= \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) =$$

$$= \left\langle \Phi(t) \equiv \frac{1}{2} + \Phi_0(t) \right\rangle =$$

$$= \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right)$$

Т. к. $X \sim N(m, \sigma^2)$ *— непрерывная случайная величина, то*

$$P\{a \preccurlyeq X \preccurlyeq b\} = \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)$$

Данные формулы важны и часто используются.

Замечание 7. *Нормальное распределение играет особую роль в теории вероятностей и математической статистике. Большинство случайных величин, описывающих естественные процессы, протекание которых зависит от большого кол-ва случайный факторов, имеет нормальное распределение.*

2.3 Случайные векторы

2.3.1 Основные понятия

Пусть

1. $(\Omega,\, \beta,\, P)$ — вероятностное пространство;
2. $X_\omega = X_1(\omega),\, \ldots,\, X_n(\omega)$ — случайные величины, заданные на этом вероятностном пространстве.

Определение. *n-мерным случайным вектором называется кортеж*[34]

$$\overrightarrow{X} = (X_1,\, \ldots,\, X_n)$$

Замечание. *В этом контексте часто используют следующую терминологию:*

1. *Случайные величины $X_1,\, \ldots,\, X_n$ называют координатами случайного вектора $\overrightarrow{X}$;*
2. *Случайный вектор $\overrightarrow{X}$ часто называют n-мерной случайной величиной.*

Пример 1. *Производят стрельбу по плоской мишени; X, Y — координаты точки попадания. Тогда $(X,\, Y)$ — двумерный случайный вектор.*

Пример 2. *У случайно выбранного пациента больницы измеряют T — температуру тела, H — рост, M — вес, P — давление, V — объём лёгких.*

Тогда $(T,\, H,\, M,\, P,\, V)$ — случайный вектор.

Замечание 1. *При рассмотрении случайных векторов мы, как правило, будем ограничиваться случаем $n = 2$.*

Замечание 2. *На реализации двумерного случайного вектора упрощённо можно смотреть как на результат случайного эксперимента, в котором на плоскость бросают точку.*

Закономерность, в соответствии с которой при многократном повторении такого эксперимента точка будет чаще или реже попадать в те или иные области на плоскости, составляет закон распределения вероятностей этого случайного вектора.

При этом задавать закон распределения случайного вектора также удобно с использованием так называемой функции распределения.

[34] Обратите внимание, что векторы обозначаются стрелочкой ($\overrightarrow{X}$), а не прямой ($\overline{X}$). Это важно, т. к. далее в курсе появится величина, которая будет обозначаться прямой. За использование прямой для обозначения вектора будут снижаться баллы. — Прим. лект.

Определение. *Функцией распределения вероятностей случайного вектора*

$$X = (X_1, \ldots, X_n)$$

называется отображение

$$F\colon \mathbb{R}^n \to \mathbb{R}$$

определённое правилом

$$F(x_1, \ldots, x_n) = P\{X_1 < x_1, \ldots, X_n < x_n\}$$

Замечание 1. *В правой части формулы, определяющей $F(x_1, \ldots, x_n)$, записана вероятность произведения событий*

$$\{X_1 < x_1\} \cdot \ldots \cdot \{X_n < x_n\}$$

Замечание 2. *В случае $n = 2$ значение $F(x_1^0, x_2^0)$ можно интерпретировать как вероятность того, что случайным образом брошенная на плоскость точка попадёт левее и ниже точки (x_1^0, x_2^0) (на диаграмме ось x — вправо, ось y — вверх).*

2.3.1.1 Свойства функции распределения случайного вектора (для n = 2)

1. $0 \leqslant F(x_1, x_2) \leqslant 1$;
2. (a) При фиксированном x_2 функция $F(x_1, x_2)$ как функция переменного x_1 является неубывающей функцией;

   (b) При фиксированном x_1 функция $F(x_1, x_2)$ как функция переменного x_2 является неубывающей функцией.
3. $\lim\limits_{x_1 \to -\infty,\, x_2 = const} F(x_1, x_2) = 0$

   $\lim\limits_{x_1 = const,\, x_2 \to -\infty} F(x_1, x_2) = 0$
4. $\lim\limits_{x_1 \to +\infty,\, x_2 \to +\infty} F(x_1, x_2) = 1$
5. $\lim\limits_{x_1 = const,\, x_2 \to +\infty} F(x_1, x_2) = F_{X_1}(x_1)$

   $\lim\limits_{x_1 \to +\infty,\, x_2 = const} F(x_1, x_2) = F_{X_2}(x_2)$

   где $F_{X_i}(x_i)$ — функция распределения случайной величины X_i;
6. Вероятность того, что реализация попадёт в похожую на прямоугольник область $D = \{(x, y)\colon x \in [a_1, b_1), y \in [a_2, b_2)\}$:

   $P\{a_1 \leqslant X < b_1, a_2 \leqslant X_2 < b_2\} = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$
7. (a) При фиксированном x_2 функция $F(x_1, x_2)$ как функция переменной x_1 является непрерывной слева в каждой точке;

(b) При фиксированном x_1 функция $F(x_1,\, x_2)$ как функция переменной x_2 является непрерывной слева в каждой точке.

Доказательства.

1. *Значение $F(x_1,\, x_2)$ является вероятностью некоторого события, следовательно, $0 \leqslant F(x_1,\, x_2) \leqslant 1$.*

2. *Доказывается аналогично одномерному случаю.*

3. *Покажем, что $\lim\limits_{x_1 \to -\infty,\, x_2 = const} F(x_1,\, x_2) = 0$.*

   По определению, $F(x_1,\, x_2) = P\{\{X_1 < x_1\} \cdot \{X_2 < x_2\}\}$; при $x_1 \to -\infty$ событие $\{X_1 < -\infty\}$ является невозможным. Произведение невозможного события на событие $\{X_2 < x_2\}$ является невозможным событием, поэтому $F(x_1,\, x_2)$ стремится к нулю при $x_1 \to -\infty$, $x_2 = const$.

   $\lim\limits_{x_1 = const,\, x_2 \to -\infty} F(x_1,\, x_2) = 0$ *доказывается аналогично.*

4. *По определению, $F(x_1,\, x_2) = P\{\{X_1 < x_1\} \cdot \{X_2 < x_2\}\}$.*

   Событие $\{X_1 < +\infty\}$ является достоверным, $\{X_2 < +\infty\}$ также является достоверным, а произведение достоверных событий — достоверное событие; таким образом,

   $$\lim_{x_1 \to +\infty,\, x_2 \to +\infty} F(x_1,\, x_2) = 1$$

5. *Покажем, что $\lim\limits_{x_1 = const,\, x_2 \to +\infty} F(x_1,\, x_2) = F_{X_1}(x_1)$*

   По определению,

   $$F(x_1,\, x_2) = P\{\{X_1 < x_1\} \cdot \{X_2 < x_2\}\}$$

   Событие $\{X_2 < +\infty\}$ является достоверным; произведение события $\{X_1 < x_1\}$ на достоверное равно $\{X_1 < x_1\}$ (т. е. равно ему же), поэтому

   $$\lim_{x_1 = const,\, x_2 \to +\infty} F(x_1,\, x_2) = P\{X_1 < x_1\} = F_{X_1}(x_1)$$

   $\lim\limits_{x_1 \to +\infty,\, x_2 = const} F(x_1,\, x_2) = F_{X_2}(x_2)$ *доказывается аналогично.*

6. (a) *Найдём вероятность попадания случайного вектора $(X_1,\, X_2)$ в полосу*

   $$\{X_1 < x_1, a_2 \leqslant X < b_2\}$$

   i. $\{X_1 < x_1,\, X_2 < b_2\} = \{X_1 < x_1,\, a_2 \leqslant X_2 < b_2\} + \{X_1 < x_1,\, X_2 < a_2\}$

ii. *По теореме сложения (события объединения несовместны):*

$$\underbrace{P\{X_1 < x_1,\, X_2 < b_2\}}_{F(x_1, b_2)} = P\{X_1 < x_1,\, a_2 \leqslant X_2 < b_2\} + \underbrace{P\{X_1 < x_1,\, X_2 < a_2\}}_{F(x_1, a_2)}$$

Таким образом,

$$P\{X_1 < x_1,\, a_2 \leqslant X_2 < b_2\} = F(x_1,\, b_2) - F(x_1,\, a_2)$$

(b) i.

$$\begin{aligned}\{X_1 < b_1,\, a_2 \leqslant X_2 < b_2\} =& \{a_1 \leqslant X_1 < b_1,\, a_2 \leqslant X_2 < b_2\} + \\ & \{X_1 < a_1,\, a_2 \leqslant X_2 < b_2\}\end{aligned}$$

ii. *По формуле сложения (события объединения несовместны):*

$$\begin{aligned}\underbrace{P\{X_1 < b_1,\, a_2 \leqslant X_2 < b_2\}}_{\text{(из пункта a) } F(b_1, b_2) - F(b_1, a_2)} =& P\{a_1 \leqslant X_1 < b_1,\, a_2 \leqslant X_2 < b_2\} + \\ & \underbrace{P\{X_1 < a_1,\, a_2 \leqslant X_2 < b_2\}}_{\text{(из пункта a) } F(a_1, b_2) - F(a_1, a_2)}\end{aligned}$$

Таким образом,

$$P\{a_1 \leqslant X_1 < b_1,\, a_2 \leqslant X_2 < b_2\} = F(b_1,\, b_2) - F(b_1,\, a_2) - F(a_1,\, b_2) + F(a_1,\, a_2)$$

7. *Доказывается аналогично одномерному случаю.*

Замечание 1. *В свойстве 5 использовались функции:* $F(x_1, x_2) = P\{X_1 < x_1,\, X_2 < x_2\}$, $F_{X_1}(x_1) = P\{X_1 < x_1\}$, $F_{X_2}(x_2) = P\{X_2 < x_2\}$. *Используется следующая терминология:* $F(x_1, x_2)$ *также называется совместной функцией распределения случайных величин* X_1 *и* X_2; $F_{X_i}(x_i)$ *называют частными (одномерными, маргинальными) функциями распределения.*

Замечание 2. *Из свойства 5 следует, что если известна функция* $F(x_1,\, x_2)$, *то всегда можно найти* $F_{X_1}(x_1) = \lim\limits_{x_2 \to +\infty} F(x_1,\, x_2)$ *и* $F_{X_2}(x_2) = \lim\limits_{x_1 \to +\infty} F(x_1,\, x_2)$.

Т. о., если известен закон распределения случайного вектора, то известны и одномерные замены распределения его координат. Верно ли обратное? Если известны $F_{X_1}(x_1)$, $F_{X_2}(x_2)$, *то можно ли найти совместную функцию распределения* $F(x_1,\, x_2)$? *Вообще говоря, нет, т. к. неизвестна связь между этими случайными величинами.*

Лекция №11, 13.11.2018

2.3.2 Дискретные случайные векторы

Определение. *Случайный вектор* $(X_1, \ldots, X_n)$ *называется дискретным, если каждая из составляющих его случайных величин* X_i, $i = \overline{1;\, n}$ *является дискретной.*

Замечание. *Очевидно, что дискретный случайный вектор может принимать лишь конечное или счётное множество значений.*

Давайте рассмотрим двумерный случайный вектор (X, Y). Для простоты будем считать, что случайные величины могут принимать значения

$$X = \{x_1, \ldots, x_m\}$$
$$Y = \{y_1, \ldots, y_n\}$$

Тогда этот вектор может принимать значения (x_i, y_j), коих $m \cdot n$ штук. Закон распределения этого вектора удобно задавать при помощи таблицы:

X, Y	y_1	$\ldots$	y_j	$\ldots$	y_n
x_1	p_{11}	$\ldots$	p_{1j}	$\ldots$	p_{1n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_i	p_{i1}	$\ldots$	p_{ij}	$\ldots$	p_{in}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_m	p_{m1}	$\ldots$	p_{mj}	$\ldots$	p_{mn}

Здесь $p_{ij} = P\{(X, Y) = (x_i, y_j)\}$, $i = \overline{1;\, m}$, $j = \overline{1;\, n}$. При этом должно выполняться условие нормировки:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1$$

Найдём вероятность того, что выпавшее значение содержит элемент x_i:

$$P\{X = x_i\} = P\{(X, Y) \in \{(x_i, y_1), \ldots, (x_i, y_n)\}\} =$$
$$= P\big\{\{(X, Y) = (x_i, y_1)\} + \ldots + \{(X, Y) = (x_i, y_n)\}\big\} =$$
$$= \langle \text{ все члены суммы несовместны } \rangle =$$
$$= \sum_{j=1}^{n} P\{(X, Y) = (x_i, y_j)\} = \underbrace{\sum_{j=1}^{n} p_{ij}}_{\text{обозначим сумму как } p_{X_i}}$$

Аналогичным образом можно найти

$$P\{Y = y_j\} = \sum_{i=1}^{m} p_{ij} = p_{Y_j}$$

Указанную выше табличку принято дополнять ещё одним столбцом для p_{X_i} и одной строкой для p_{Y_j}:

X, Y	y_1	$\dots$	y_j	$\dots$	y_n	P_X
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1n}	P_{X_1}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$	p_{in}	P_{X_i}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	p_{m1}	$\dots$	p_{mj}	$\dots$	p_{mn}	P_{X_m}
P_Y	P_{Y_1}	$\dots$	P_{Y_j}	$\dots$	P_{Y_n}	1

Сумма элементов в каждой строке исходной таблицы должна равняться соответствующему P_{X_i}, а сумма элементов в каждом столбце — P_{Y_j}. Сумма элементов в строке P_Y (и сумма элементов в столбце P_X, отдельно) должна равняться 1.

Пример. *Два раза подбрасывают симметричную монету: X — кол-во выпадений герба, а Y — номер броска, при котором герб выпал впервые ($Y = 3$, если герб ни разу не выпал). Найдём закон распределения вектора (X, Y).*

$$X \in \{0, 1, 2\},\ Y \in \{1, 2, 3\}$$

Тогда

X, Y	1	2	3	P_X
0	0	0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
2	$\frac{1}{4}$	0	0	$\frac{1}{4}$
P_Y	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1

2.3.3 Непрерывные случайные векторы

Определение. *Случайный вектор $(X_1, \dots, X_n)$ называется непрерывным, если существует функция*

$$f\colon \mathbb{R}^n \to \mathbb{R}$$

такая, что для каждой точки $(x_1, \dots, x_n)$ выполняется

$$F(x_1, \dots, x_n) = \int\limits_{-\infty}^{x_1} dt_1 \int\limits_{-\infty}^{x_2} dt_2 \dots \int\limits_{-\infty}^{x_i} dt_i \dots \int\limits_{-\infty}^{x_n} f(t_1, \dots, t_n)\, dt_n$$

где F — функция распределения плотности случайного вектора $(X_1, \dots, X_n)$. При этом f называется функцией плотности распределения вероятностей этого вектора.

Замечание 1. *В определении предполагается. что несобственный интеграл сходится в каждой точке $(x_1, \dots, x_n) \in \mathbb{R}^n$.*

Замечание 2. *При $n = 2$:*

$$F(x_1,\, x_2) = \int\limits_{-\infty}^{x_1} dt_1 \int\limits_{-\infty}^{x_2} f(t_1,\, t_2)\, dt_2$$

Замечание 3. *Всюду в дальнейшем будем предполагать, что функция плотности $f(x_1,\, \ldots,\, x_n)$ непрерывна всюду, кроме, быть может, множеств меры нуль. Для $n = 2$ это означает, что функция плотности $f(x_1,\, x_2)$ непрерывна на всей плоскости, кроме, быть может, отдельных точек или линий.*

Замечание 4. *С использованием теоремы о производной интеграла с переменным верхним пределом получаем, что если $(x_1^0,\, \ldots,\, x_n^0)$ — точка непрерывности функции F, то*

$$f(x_1^0,\, \ldots,\, x_n^0) = \left.\frac{\delta^n F(x_1,\, \ldots,\, x_n)}{\delta x_1\, \delta x_2\, \ldots\, \delta x_n}\right|_{(x_1^0,\, \ldots,\, x_n^0)} \tag{11}$$

Для $n = 2$:

$$f(x_1^0,\, x_2^0) = \left.\frac{\delta^2 F(x_1,\, x_2)}{\delta x_1\, \delta x_2}\right|_{(x_1^0,\, x_2^0)}$$

Получается, что из функции плотности можно получить функцию распределения случайного вектора (см. определение непрерывной случайной величины), и, наоборот, из функции распределения можно получить функцию плотности (см. 11).

Таком образом, функция плотности, как и функция распределения случайного вектора, содержит всю информацию о его законе распределения. Поэтому задавать закон распределения случайного вектора можно как с использованием функции плотности, так и с использованием функции распределения.

2.3.3.1 Свойства непрерывных случайных векторов (для n=2)

1. Если f — функция плотности двумерного случайного вектора, то $f(x_1,\, x_2) \geqslant 0$.

2. Если $(X_1,\, X_2)$ — непрерывный случайный вектор, а $f(x_1,\, x_2)$ — его функция плотности, то

$$P\{a_1 \preccurlyeq X_1 \preccurlyeq b_1,\, a_2 \preccurlyeq X_2 \preccurlyeq b_2\} = \int\limits_{a_1}^{b_1} dx_1 \int\limits_{a_2}^{b_2} f(x_1,\, x_2)\, dx_2$$

3. Условие нормировки: $\iint\limits_{R^2} f(x_1,\, x_2)\, dx_1\, dx_2 = 1$.

4. Если $f(x_1,\, x_2)$ — функция плотности вектора (X_1, X_2), а (x_1^0, x_2^0) — точка непрерывности функции f, то

$$P\{x_1^0 \preccurlyeq X_1 \preccurlyeq x_1^0 + \triangle x_1, x_2^0 \preccurlyeq X_2 \preccurlyeq x_2^0 + \triangle x_2\} \approx f(x_1^0,\, x_2^0) \triangle x_1 \triangle x_2$$

если $\triangle x_1$, $\triangle x_2$ достаточно малы.

5. Если $(X_1,\, X_2)$ — непрерывный случайный вектор, то для любых наперёд заданных x_1^0, x_2^0

$$P\{(X_1,\, X_2) = (x_1^0,\, x_2^0)\} = 0$$

6.

$$P\{(X,\, Y) \in D\} = \iint\limits_D f(x_1,\, x_2)\, dx_1\, dx_2$$

7.

$$\int\limits_{-\infty}^{+\infty} f(x_1,\, x_2)\, dx_2 = f_{X_1}(x_1)$$

$$\int\limits_{-\infty}^{+\infty} f(x_1,\, x_2)\, dx_1 = f_{X_2}(x_2)$$

где f_{X_1}, f_{X_2} — маргинальные функции плотности случайных величин X_1 и X_2, f — совместная функция плотности случайных величин X_1 и X_2 ($\equiv$ функция плотности случайного вектора $(X_1,\, X_2)$).

Обратите внимание, что из функции плотности можно получить обе маргинальные.

$$f(x_1,\, x_2) \implies \begin{cases} f_{X_1}(x_1) \\ f_{X_2}(x_2) \end{cases}$$

Доказательства. *Доказательства свойств 1-5 аналогичны одномерному случаю. Свойство 6 является обобщением свойства 2 на случай произвольной области D (без доказательства).*

Доказательство свойства 7.

Докажем, что $f_{X_1}(x_1) = \int\limits_{-\infty}^{+\infty} f(x_1,\, x_2)\, dx_2$.

По свойству двумерной функции распределения $F(x_1,\, +\infty) = F_{X_1}(x_1)$*; таким образом (подставим определение функции распределения для двумерного вектора),*

$$F_{X_1}(x_1) = \int\limits_{-\infty}^{x_1} dt_1 \int\limits_{-\infty}^{+\infty} f(t_1,\, t_2)\, dt_2$$

$$f_{X_1}(x_1) = \frac{dF_{X_1}(x_1)}{dx_1} =$$

$$= \Big\langle x_1 \text{ — точка непрерывности функции } f_{X_1}(x_1), \text{ и тогда по теореме}$$

$$\text{о производной интеграла с перменным верхним пределом} \Big\rangle =$$

$$= \int\limits_{-\infty}^{+\infty} f(x_1,\, t_2)\, dt_2$$

Вторая формула доказывается аналогично.

Лекция №12, 20.11.2018

Пример. *Функция плотности распределения вероятностей случайного вектора* $(X,\, Y)$ *имеет вид*

$$f(x,\, y) = \begin{cases} cxy, & (x,\, y) \in K, \\ 0 & \text{иначе} \end{cases}$$

где K *— квадрат стороной* 1 *(см. рисунок 29).*

Нужно найти:

1. *Постоянную c;*
2. *Маргинальные плотности распределения* $f_X(x)$, $f_Y(y)$.

Решение.

1. *Условие нормировки:*

$$1 = \iint\limits_{R^2} f(x,\, y)\, dx\, dy =$$

$$= \Big\langle f(x,\, y) \equiv 0 \text{ при } (x,\, y) \notin K \Big\rangle =$$

$$= \iint\limits_{K} cxy\, dx\, dy = c \int\limits_0^1 x\, dx \int\limits_0^1 y\, dy = c \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{c}{4}$$

Таким образом, $1 = \frac{c}{4} \implies c = 4$. *Итого:*

$$f(x,\, y) = \begin{cases} 4xy, & (x,\, y) \in K, \\ 0 & \textit{иначе} \end{cases}$$

2. (a) *Найдём* $f_X(x)$.
   Используя свойство 7

$$f_X(x) = \int\limits_{-\infty}^{+\infty} f(x,\, y)\, dy = \begin{cases} \int_0^1 4xy\, dy, & x \in (0,\, 1) \\ 0, & x \notin (0,\, 1) \end{cases} = \begin{cases} 2x, & x \in (0,\, 1) \\ 0, & x \notin (0,\, 1) \end{cases}$$

   (b) *Аналогично можно найти*

$$f_Y(y) = \ldots = \begin{cases} 2y, & y \in (0,\, 1) \\ 0, & y \notin (0,\, 1) \end{cases}$$

2.3.4 Независимая случайная величина

Давайте рассмотрим двумерный случайный вектор $(X,\, Y)$, множество возможных значений которого конечно. Пусть

$$X \in \{x_1,\, \ldots,\, x_n\},\; Y \in \{y_1,\, \ldots,\, y_n\}$$

По аналогии с определением независимых событий определение независимых случайных величин X и Y в рассматриваемом случае можно определить в следующем виде:

Нестрогое определение. X, Y *— независимы, если*

$$P\{(X,\, Y) = (x_i,\, y_j)\} \equiv P\{X = x_i\}\, P\{Y = y_j\},\; i = \overline{1;\, m},\, j = \overline{1;\, l} \tag{12}$$

Посмотрим, что можно в этом случае можно сказать о совместной функции распределения случайных величин X и Y:

$$\begin{gathered} F(x,\, y) = P\{X < x,\, Y < y\} = \langle (X,\, Y) - \text{дискретный случайный вектор } \rangle = \\ = P\{X \in \{x_1,\, \ldots,\, x_n\},\, Y \in \{y_1,\, \ldots,\, y_n\}\} = \\ = P\{(X,\, Y) \in \{x_i,\, y_j\} \colon i = \overline{1;\, k},\, j = \overline{1;\, p}\} = \\ = \langle \text{ события } \{(x_i,\, y_j)\} \text{ для различных } (i,\, j) \text{ несовместны } \rangle = \\ = \sum_{i=1}^{k} \sum_{j=1}^{p} P\{(X,\, Y) = (x_i,\, y_j)\} = \langle \text{ см. предварительное определение } \rangle = \\ = \sum_{i=1}^{k} \sum_{j=1}^{p} P\{X = x_i\} \cdot P\{Y = y_j\} = \left[\sum_{i=1}^{k} P\{X = x_i\}\right] \cdot \left[\sum_{j=1}^{p} P\{Y = y_j\}\right] = \\ = P\{X < x\}\, P\{Y < y\} = F_X(x)\, F_Y(y) \end{gathered}$$

Таким образом, для произвольных случайных величин X и Y дадим следующее:

Определение. *Случайные величины X и Y называются независимыми, если*

$$F(x,\, y) = F_X(x)\, F_Y(y)$$

где F — совместная функция распределения X и Y ($\equiv$ функция распределения случайного вектора $(X,\, Y)$); $F_X,\, F_Y$ — маргинальные функции распределения случайных величин X и Y.

2.3.4.1 Свойства независимых случайных величин

1.
$$\text{Случайные величины } X \text{ и } Y \text{ независимы}$$
$$\Longleftrightarrow$$
$$\forall\forall\, x,\, y \in \mathbb{R} \text{ события } \{X < x\} \text{ и } \{Y < y\} \text{ независимы.}$$

2.
$$\text{Случайные величины } X \text{ и } Y \text{ независимы}$$
$$\Longleftrightarrow$$
$$\forall\forall x_1,\, x_2 \in \mathbb{R},\, \forall\forall y_1,\, y_2 \in \mathbb{R} \text{ события } \{x_1 \leqslant X < x_2\} \text{ и } \{y_1 \leqslant Y < y_2\} \text{ независимы.}$$

3.
$$\text{Случайные величины } X \text{ и } Y \text{ независимы}$$
$$\Longleftrightarrow$$
$$\forall\forall M_1,\, M_2 \text{ события } \{X \in M_1\} \text{ и } \{Y \in M_2\} \text{ независимы,}$$

   где $M_1,\, M_2$ — промежутки или объединения промежутков в $\mathbb{R}$.

4. Если X и Y — дискретные случайные величины, то

$$X,\, Y \text{ — независимые } \iff p_{ij} \equiv P_{X_i} P_{Y_j}$$

   где $p_{ij} = P\{(X,\, Y) = (x_i,\, y_j)\}$, $P_{X_i} = P\{X = x_i\}$, $P_{Y_j} = P\{Y = y_j\}$.

5. Если X и Y — непрерывные случайные величины, то

$$X,\, Y \text{ — независимы } \iff f(x,\, y) \equiv f_X(x)\, f_Y(y)$$

   где f — совместная плотность распределения случайных величин X и Y ($\equiv$ функция плотности распределения случайного вектора $(X,\, Y)$); $f_X,\, f_Y$ — маргинальные плотности распределения случайных величин X и Y соответственно.

Доказательства.

1. *Очевидно следует из определения независимых случайных величин.*

2. (a) ***Необходимость ($\Longrightarrow$).***
Пусть $F(x,\, y) = F_X(x)\, F_Y(y)$.
Тогда

$$\begin{gathered}P\{x_1 \leqslant X < x_2,\, y_1 \leqslant Y < y_2\} = \\ = \langle\ \text{свойство функции распределения случайного вектора}\ \rangle = \\ = F(x_1,\, y_1) + F(x_2,\, y_2) - F(x_1,\, y_2) - F(x_2,\, y_1) = \\ = F_X(x_1)\, F_Y(y_1) + F_X(x_2)\, F_Y(y_2) - F_X(x_1)\, F_Y(y_2) - F_X(x_2)\, F_Y(y_1) = \\ = [F_X(x_2) - F_X(x_1)]\, [F_Y(y_2) - F_Y(y_1)] = \\ = \langle\ \text{свойство одномерной функции распределения}\ \rangle = \\ = P\{x_1 \leqslant X < x_2\}\, P\{y_1 \leqslant Y < y_2\}\end{gathered}$$

   (b) ***Достаточность ($\Longleftarrow$).***
Пусть $\forall\forall x_1,\, x_2,\, y_1,\, y_2 \in \mathbb{R}$,

$$P\{x_1 \leqslant X < x_2,\, y_1 \leqslant Y < y_2\} = P\{x_1 \leqslant X < x_2\}\, P\{y_1 \leqslant Y < y_2\}$$

Тогда

$$\begin{gathered}F(x,\, y) = P\{X < x,\, Y < y\} = P\{-\infty < X < x,\, -\infty < Y < y\} = \\ = \langle x_1 = -\infty,\, x_2 = x,\ y_1 = -\infty,\, y_2 = y \rangle = \\ = P\{-\infty < X < x\}\, P\{-\infty < Y < y\} = F_X(x)\, F_Y(y)\end{gathered}$$

3. *Является обобщением свойств 1 и 2 (без доказательства).*

4. (a) ***Достаточность ($\Longleftarrow$).*** *Достаточность была доказана выше, в рассуждениях перед определением независимых случайных величин.*

   (b) ***Необходимость ($\Longrightarrow$).*** *Необходимость студентам предлагается доказать самостоятельно.*

5. (a) ***Необходимость ($\Longrightarrow$).***
Пусть $F(x,\, y) \equiv F_X(x)\, F_Y(y)$. *По свойству двумерной плоскости*

$$f(x,\, y) = \frac{\delta^2 F(x,\, y)}{\delta x\, \delta y} = \frac{\delta^2}{\delta x\, \delta y}\, [F_X(x)\, F_Y(y)] = \frac{dF_X(x)}{dx} \cdot \frac{dF_Y(y)}{dy} = f_X(x)\, f_Y(y)$$

   (b) ***Достаточность ($\Longleftarrow$).***
Пусть $f(x,\, y) = f_X(x)\, f_Y(y)$.

Тогда

$$F(x,\, y) = \int\limits_{-\infty}^{x} dt \int\limits_{-\infty}^{y} f(t,\, v)\, dv = \int\limits_{-\infty}^{x} dt \int\limits_{-\infty}^{y} f_X(t)\, f_Y(v)\, dv =$$

$$\underbrace{\int\limits_{-\infty}^{x} f_X(t)\, dt}_{F_X(x)} \underbrace{\int\limits_{-\infty}^{y} f_Y(v) dv}_{F_Y(y)} = F_X(x)\, F_Y(y)$$

Пример 1. *Рассмотрим дискретный случайный вектор из примера о подбрасывании монеты:*

X, Y	1	2	3	P_X
0	0	0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
2	$\frac{1}{4}$	0	0	$\frac{1}{4}$
P_Y	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Воспользуемся свойством 4 (X, Y — независимы $\iff p_{ij} \equiv P_{X_i} P_{Y_j}$):

$$p_{11} = 0 \neq \frac{1}{4} \cdot \frac{1}{2} = P_{X_1} P_{Y_1} \implies X,\, Y \text{ — зависимые.}$$

Пример 2. *Рассмотрим непрерывный случайный вектор из примера про квадрат:*

$$f(x,\, y) = \begin{cases} 4xy, & (x,\, y) \in K, \\ 0 & \text{иначе} \end{cases}$$

Были найдены две маргинальных компоненты этого вектора

$$f_X(x) = \begin{cases} 2x, & x \in (0,\, 1) \\ 0, & x \notin (0,\, 1) \end{cases} \qquad f_Y(y) = \begin{cases} 2y, & y \in (0,\, 1) \\ 0, & y \notin (0,\, 1) \end{cases}$$

Т. к. $f(x,\, y) \equiv f_X(x)\, f_Y(y)$, то X, Y — независимые (по свойству 5).

Определение. *Случайные величины $X_1, \ldots, X_n$, заданные на одном вероятностном пространстве, называются*

- *Попарно независимыми, если X_i и X_j независимы при $i \neq j$;*

- *Независимыми в совокупности, если*

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$$

где F — совместная функция распределения случайных величин $X_1, \ldots X_n$ ($\equiv$ функция распределения случайного вектора $(X_1, \ldots X_n)$), F_{X_i} — маргинальные функции распределения случайных величин X_i, $i = \overline{1; n}$.

Замечание 1. *Можно доказать, что*

1. *Если $X_1, \ldots, X_n$ независимы в совокупности, то они попарно независимы. Обратное неверно.*

2. *Обобщения свойств 4 и 5 будут справедливы для любого числа n случайных величин, независимых в совокупности. К примеру, обобщение свойства 5:*

$$X_1, \ldots, X_n \text{ — независимы в совокупности}$$
$$\Longleftrightarrow$$
$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

Лекция №13, 27.11.2018

2.3.5 Условные распределения

Давайте рассмотрим случайный вектор (X, Y). Предположим, известно, что случайная величина Y приняла значение y_0. Что в этом случае можно сказать о возможных значениях случайной величины X и что можно сказать о законе распределения случайной величины X при условии $Y = y_0$?

2.3.5.1 Случай дискретного случайного вектора

Пусть

1. (X, Y) — дискретный случайный вектор;

2. $X \in \{x_1, \ldots, x_m\}$, $Y \in \{y_1, \ldots, y_n\}$;

3. $p_{ij} = P\{(X, Y) = (x_i, y_j)\},\ i = \overline{1; m},\ j = \overline{1; n}$,
   $P_{X_i} = P\{X = x_i\},\ i = \overline{1; m}$,
   $P_{Y_j} = P\{Y = y_j\},\ j = \overline{1; n}$;

4. Известно, что $Y = y_j$ для некоторого фиксированного j.

Тогда

$$P\{X = x_i \,|\, Y = y_j\} = \langle \text{ из определения условной вероятности } \rangle =$$
$$= \frac{P\{\{X = x_i\} \cdot \{Y = y_j\}\}}{P\{Y = y_j\}} = \frac{P\{(X,\, Y) = (x_i,\, y_j)\}}{P_{Y_j}} = \frac{p_{ij}}{P_{Y_j}}$$

Определение. *В случае двумерного дискретного случайного вектора $(X,\, Y)$ условной вероятностью того, что случайная величина X приняла значение x_i при условии $Y = y_j$ называют число*

$$\pi_{ij} = \frac{p_{ij}}{P_{Y_j}}$$

Определение. *Набор вероятностей π_{ij}, $i = \overline{1;\, m}$, для данного фиксированного j называется условным распределением случайной величины X при условии $Y = y_j$.*

Замечание 1. *Условная вероятность того, что случайная величина Y приняла значение y_j при условии $X = x_i$ определяется аналогично*[35]*:*

$$\tau_{ij} = P\{Y = y_j \,|\, X = x_i\} = \frac{p_{ij}}{P_{X_i}}$$

Замечание 2. *Набор вероятностей τ_{ij}, $j = \overline{1;\, n}$, для фиксированного i называют условным распределением случайной величины Y при условии $X = x_i$.*

Пример. *Рассмотрим двумерный случайный вектор $(X,\, Y)$ из задачи о подбрасывании монеты.*

X, Y	1	2	3	P_X
0	0	0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
2	$\frac{1}{4}$	0	0	$\frac{1}{4}$
P_Y	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Из этого можно получить ряд распределения случайной величины X:

x	0	1	2
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

А также ряд распределения случайной величины Y:

y	1	2	3
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

[35] τ — строчная буква «тау» греческого алфавита. — Прим. ред.

Можно составить таблицу условных распределений по π_{ij}:

X, Y	1	2	3
0	0	0	1
1	$\frac{1}{2}$	1	0
2	$\frac{1}{2}$	0	0
	1	1	1

В такой таблице каждый столбец соответствует условным распределениям случайной величины X при условиях, последовательно, $Y = 1$, $Y = 2$, $Y = 3$. В первом столбце находятся значения π_{i1}, $i = \overline{1;\,3}$, во втором — π_{i2}, $i = \overline{1;\,3}$, третьем — π_{i3}, $i = \overline{1;\,3}$.

Аналогичную таблицу можно сделать для τ_{ij}.

2.3.5.2 Случай непрерывного случайного вектора

В случае непрерывного случайного вектора (X, Y) рассуждения, аналогичные проведённым для дискретного случайного вектора рассуждениям, приводят к следующему определению.

Определение. *Условной плотностью распределения случайного вектора X при условии $Y = y$ называется функция*

$$f_X(x \,|\, Y = y) = \frac{f(x,\, y)}{f_Y(y)}$$

где f — совместная плотность распределения случайных величин X и Y ($\equiv$ плотность распределения вектора (X, Y)), f_Y — маргинальная плотность распределения случайной величины Y.

Замечание. *Аналогичным образом определяется условная плотность распределения случайной величины Y при условии $X = x$:*

$$f_Y(y \,|\, X = x) = \frac{f(x,\, y)}{f_X(x)}$$

где $f_X(x)$ — маргинальная плотность распределения случайной величины X.

Пример. *Случайный вектор (X, Y) распределён равномерно в круге K радиуса R с центром в начале координат. Найти условные законы распределения компонент этого вектора.*

Решение. *(X, Y) распределены равномерно в K, следовательно*

$$f(x,\, y) = \begin{cases} c, & (x,\, y) \in K \\ 0, & (x,\, y) \notin K \end{cases}$$

Константа c выражается из условия нормировки:

$$1 = \iint\limits_{R^2} f(x,\, y)\, dx\, dy = \langle f(x,\, y) \equiv 0 \text{ вне } K \rangle =$$

$$= \iint\limits_{K} c\, dx\, dy = c \cdot (\text{площадь } K) = c \cdot \pi R^2 = 1 \implies c = \frac{1}{\pi R^2}$$

Таким образом,

$$f(x,\, y) = \begin{cases} \dfrac{1}{\pi R^2}, & (x,\, y) \in K \\ 0 & \text{иначе} \end{cases}$$

Для того, чтобы найти условные плотности, найдём маргинальные плотности:

$$f_X(x) = \int\limits_{-\infty}^{+\infty} f(x,\, y)\, dy = \begin{cases} 0, & \text{если } |x| > R \\ \int\limits_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \dfrac{1}{\pi R^2}\, dy, & \text{если} |x| < R \end{cases} = \begin{cases} \dfrac{2\sqrt{R^2 - x^2}}{\pi R^2}, & |x| < R \\ 0, & |x| > R \end{cases}$$

Аналогично:

$$f_X(x) = \begin{cases} \dfrac{2\sqrt{R^2 - y^2}}{\pi R^2}, & |y| < R \\ 0, & |y| > R \end{cases}$$

Найдём условные плотности:

$$f_X(x \,|\, Y = y) = \frac{f(x,\, y)}{f_Y(y)} = \begin{cases} \text{не определена}, & |y| > R \\ 0, & |y| < R,\, x^2 + y^2 > R^2 \\ \dfrac{\left(\dfrac{1}{\pi R^2}\right)}{\left(\dfrac{2\sqrt{R^2 - y^2}}{\pi R^2}\right)}, & |y| < R, x^2 + y^2 < R^2 \end{cases} =$$

$$= \begin{cases} \dfrac{1}{2\sqrt{R^2 - y^2}}, & |y| < R,\, x^2 + y^2 < R^2 \\ 0, & |y| < R,\, x^2 + y^2 > R^2 \\ \text{не определена}, & |y| > R \end{cases}$$

Аналогично (в силу симметрии):

$$f_Y(y \,|\, X = x) = \begin{cases} \dfrac{1}{2\sqrt{R^2 - x^2}}, & |x| < R,\ x^2 + y^2 < R^2 \\ 0, & |x| < R,\ x^2 + y^2 > R^2 \\ \textit{не определена}, & |x| > R \end{cases}$$

Теорема. ***Критерии независимости случайных величин в терминах условных распределений.***

1. *Пусть (X, Y) — двумерный случайный вектор. Тогда следующие условия эквивалентны:*
   - *X, Y — независимые.*
   - *$F_X(x \,|\, Y = y) \equiv F_X(x)$ для всех y, в которых определена $F_X(x \,|\, Y = y)$.*
   - *$F_Y(y \,|\, X = x) \equiv F_Y(y)$ для всех x, в которых определена $F_Y(y \,|\, X = x)$.*
2. *Если (X, Y) — непрерывный случайный вектор, то следующие условия эквивалентны. (для условной плотности)*
   - *X, Y — независимые.*
   - *$f_X(x \,|\, Y = y) \equiv f_X(x)$ для всех y, в которых определена $f_X(x|Y = y)$.*
   - *$f_Y(y \,|\, X = x) \equiv f_Y(y)$ для всех x, в которых определена $f_Y(y|X = x)$.*
3. *Если (X, Y) — дискретный случайный вектор, то следующие утверждения эквиваленты.*
   - *X, Y — независимые.*
   - *$P\{X = x_i \,|\, Y = y_j\} \equiv P\{X = x_i\}$ для всех $j = \overline{1;\, n}$.*
   - *$P\{Y = y_j \,|\, X = x_i\} \equiv P\{Y = y_j\}$ для всех $i = \overline{1;\, m}$.*

   (здесь $X \in \{x_1, \ldots, x_m\}$, $Y \in \{y_1, \ldots, y_n\}$).

Доказательство. *Без доказательства.*

Замечание. *Пусть (X, Y) — случайный вектор.*

Определение. *Условной функцией распределения случайной величины X при условии $Y = y$ называют функцию*

$$F_X(x \,|\, Y = y) = P\{X < x \,|\, Y = y\}$$

Аналогично:

$$F_Y(y \,|\, X = x) = P\{Y < y \,|\, X = x\}$$

Можно доказать, что условная функция распределения при фиксированном условии обладают всеми свойствами обычной функции распределения.

Пример 1. *Рассмотрим дискретный случайный вектор из задачи о подбрасывании монеты. Ранее мы вычислили условное распределение* π_{ij}. *Например,* $P\{X = 0 \,|\, Y = 2\} = 1$, *но* $P\{X = 0\} = \frac{1}{4}$. $1 \neq \frac{1}{4} \implies X$ *и* Y *зависимы.*

Пример 2. *Рассмотрим случайный вектор* (X, Y), *распределённый равномерно в круге (см. пример выше).*

$$f_X(x \,|\, Y = y) = \begin{cases} \dfrac{1}{2\sqrt{R^2 - y^2}}, & |y| < R,\, x^2 + y^2 < R^2 \\ 0, & |y| < R,\, x^2 + y^2 > R^2 \\ \textit{не определена}, & |y| > R \end{cases}$$

$$f_X(x) = \begin{cases} \dfrac{2\sqrt{R^2 - x^2}}{\pi R^2}, & |x| < R \\ 0, & |x| > R \end{cases}$$

Т. к. $f_X(x \,|\, Y = y) \neq f_X(x)$ *(например, при* $y = 0$*), то* X, Y *— зависимые.*

2.4 Функции от случайных величин

2.4.1 Скалярная функция от одномерной случайной величины

Пусть

1. X — некоторая случайная величина;
2. $\varphi\colon \mathbb{R} \to \mathbb{R}$ — некоторая известная функция.

Тогда $\varphi(X) = Y$ — некоторая случайная величина.

Пример. *Пусть X — радиус шара — случайная величина. Тогда объём шара $Y = \frac{4}{3}\pi X^3$ — тоже случайная величина. Здесь $\varphi(x) = \frac{4}{3}\pi x^3$.*

Основной вопрос. Как, зная закон распределения случайной величины X и функцию φ, найти закон распределения случайной величины $Y = \varphi(X)$?

2.4.1.1 Случай дискретной случайной величины

Пусть X — дискретная случайная величина, имеющая ряд распределения

$$\begin{array}{cccccc} X & x_1 & \dots & x_i & \dots & x_n \\ P & p_1 & \dots & p_i & \dots & p_n \end{array}$$

Тогда $Y = \varphi(X)$ — тоже дискретная случайная величина. При этом Y принимает значения $\varphi(x_1), \dots, \varphi(x_n)$. Поэтому ряд распределения Y будет выглядеть вот так:

$$\begin{array}{cccccc} Y & \varphi(x_1) & \dots & \varphi(x_i) & \dots & \varphi(x_n) \\ P & p_1 & \dots & p_i & \dots & p_n \end{array}$$

Если в этой таблице некоторые из значений $\varphi(x_i)$ совпадают, то соответствующие столбцы нужно объединить, приписав этому значению суммарную вероятность.

Пример. *Пусть*

$$\begin{array}{cccc} X & -1 & 0 & 1 \\ P & 0.2 & 0.7 & 0.1 \end{array}$$

Пусть $Y = X^2 + 1$, т. е. $\varphi(x) = x^2 + 1$.
Найдём ряд распределения случайной величины Y:

$$\begin{array}{cccc} Y & 2 & 0 & 2 \\ P & 0.2 & 0.7 & 0.1 \end{array}$$

из чего следует

$$\begin{array}{ccc} Y & 0 & 2 \\ P & 0.7 & 0.3 \end{array}$$

Лекция №14, 04.12.2018

2.4.1.2 Случай непрерывной случайной величины

Если X — непрерывная случайная величина, то в зависимости от функции φ случайная величина $Y = \varphi(X)$ может быть как непрерывно случайной величиной, так и дискретной или смешанного типа.

Теорема. *Пусть*

1. *X — непрерывная случайная величина;*

2. *$\varphi\colon \mathbb{R} \to \mathbb{R}$;*

3. *φ монотонна и непрерывно дифференцируема;*

4. *ψ — функция[36], обратная к φ (т. к. φ — монотонная, то $\exists \psi = \varphi^{-1}$);*

5. *$Y = \varphi(X)$.*

Тогда

1. *Y также является непрерывной случайной величиной;*

2. *$f_Y(y) = f_X(\psi(y))|\psi'(y)|$.*

Доказательство.

1. $F_Y(y) = P\{Y < y\} = P\{\varphi(X) < y\}$

   (a) *Если φ — монотонно возрастающая функция, то $\varphi(X) < y \iff X < \varphi^{-1}(y) = \psi(y)$;*

   (b) *Если φ — монотонная убывающая функция, то $\varphi(X) < y \iff X > \varphi^{-1}(y) = \psi(y)$*

2. *В случае случая а, $F_Y(y) = P\{X < \psi(y)\} = F_X(\psi(y))$; в случае б $F_Y(y) = P\{X > \varphi(y)\} = 1 - P\{X \preccurlyeq \psi(y)\} = \langle$ X — непрерывная $\rangle = 1 - P\{X < \psi(y)\} = 1 - F_X(\psi(y))$.*

3.

$$f_Y(y) = F'_Y(y) = \begin{cases} \dfrac{d}{dy}[F_X(\psi(y))] = F'_X(\psi(y)) \cdot \psi'(y), & \text{если а} \\ \dfrac{d}{dy}[1 - F_X(\psi(y))] = -F'_X(\psi(y)) \cdot \psi'(y), & \text{если б} \end{cases}$$
$$= f_X(\psi(y)) \cdot |\psi'(y)|$$

[36] ψ — строчная буква «пси» греческого алфавита. — Прим. ред.

Пример 1. *Пусть $X \sim Exp(\lambda)$. Найдём закон распределения случайной величины $Y = e^X$.*

Решение.

1. $X \sim Exp(\lambda) \implies f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$

2. $Y = e^X$, *т. е.* $\varphi(x) = y \iff y = \ln x$, *т. е.* $\psi(y) = \ln y$

3. *Найдём* $f_Y(y)$:

$$f_Y(y) = f_X(\psi(y))\,|\psi'(y)| =$$

$$= \left\langle \textit{т. к. } Y = e^X, \textit{ то } Y \geqslant 0 \implies f_Y(y) \equiv 0,\, y < 0 \right\rangle =$$

$$= \begin{cases} 0, & y < 0 \\ f_X(\ln y)\,|(\ln y)'_y|, & y > 0 \end{cases}$$

$$= \begin{cases} 0, & y < 0 \\ 0, & \ln y < 0 \ \langle \implies \textit{при } y < 1\ f_Y = 0 \rangle \\ \lambda e^{-\lambda \ln y} \left|\dfrac{1}{y}\right|, & \ln y > 0 \end{cases}$$

$$= \begin{cases} \dfrac{\lambda e^{\ln y^{-\lambda}}}{y}, & y > 1 \\ 0, & y < 1 \end{cases}$$

$$= \begin{cases} \lambda y^{-\lambda - 1}, & y > 1 \\ 0, & y < 1 \end{cases}$$

Пример 2. *Пусть*

1. X *— непрерывная случайная величина, которая имеет функцию распределения* $F(x)$;

2. $F(x)$ *непрерывна;*

3. $Y = F(x)$, *т. е.* $\varphi = F$.

Найдём закон распределения случайной величины Y.

Решение. *Очевидно, что* $Y \in [0,\,1]$.

Если $y \leqslant 0$, *то* $F_Y(y) = 0$.

Если $0 < y \leqslant 1$, *то*

$$F_Y(y) = \langle \textit{ см. доказательство предыдущей теоремы, пункт а } \rangle$$
$$= F(\underbrace{\varphi^{-1}}_{F^{-1}}(y)) = F(F^{-1}(y)) = y$$

Если $y > 1$, то $F_Y(y) = 1$.

Таким образом, $F_Y(y) = \begin{cases} 0, & y \leqslant 0 \\ y, & 0 < y \leqslant 1 \\ 1, & 1 < y \end{cases} \implies f_Y(y) = \begin{cases} 1, & y \in (0,\,1), \\ 0 & \text{иначе} \end{cases}$

Таким образом, $Y \sim R(0;\,1)$ (т. е. Y равномерно распределена на $(0;\,1)$).

Замечание. *Из этого примера следует, что если Y — равномерно распределённое на $(0;\,1)$ случайная величина, то $X = F^{-1}(Y)$ будет иметь F своей функцией распределения. Этот результат широко используется при компьютерном моделировании случайных величин. Достаточно иметь генератор случайных чисел в интервале $(0;\,1)$ (с равномерным распределением), а для получения реализаций случайной величины X с функцией распределения $F(x)$ нужно сгенерированное из $(0;\,1)$ значение подвергнуть функциональному преобразованию функцией F^{-1}.*

Теорема. *Пусть*

1. *X — непрерывная случайная величина;*

2. *$\varphi\colon \mathbb{R} \to \mathbb{R}$ является кусочно-монотонной функцией, имеющей n интервалов монотонности;*

3. *φ дифференцируема;*

4. *Для данного $y \in R$, $x_1 = x_1(y), \ldots, x_k = x_k(y)$ $(k \leqslant n)$ — это все решения уравнения $y = \varphi(x)$, принадлежащие интервалам $I_1, \ldots, I_k$ монотонности функции φ.*

   Тогда для данного в условии 4 значения y

$$f_Y(y) = \sum_{j=1}^{k} f_X(\psi_j(y)) * |\psi_j'(y)|$$

где $\psi_j(y)$ — функция, обратная к $\varphi(x)$ на интервале I_j, $j = \overline{1;\,k}$

Доказательство. *Без доказательства.*

2.4.2 Скалярные функции случайного вектора

Пусть

1. $(X_1,\,X_2)$ — двумерный случайный вектор;

2. $\varphi : \mathbb{R}^2 \to \mathbb{R}$

3. $Y = \varphi(X_1,\,X_2)$ — некоторая одномерная случайная величина.

Как, зная закон распределения случайного вектора (X_1, X_2), найти закон распределения случайной величины Y?

Пример. *Пусть (X_1, X_2) — координаты попадания пули при стрельбе по плоской мишени. $Y = \sqrt{X_1^2 + X_2^2}$ — расстояние от точки попадания пули до центра мишени.*

Рассмотрим два случая.

2.4.2.1 Случай дискретного случайного вектора

Пусть (X_1, X_2) — дискретный случайный вектор. В таком случае Y — дискретная случайная величина.

Пример. *Проводится два испытания по схеме Бернулли с вероятностью успеха p:*

$$X_1 = \begin{cases} 1, & \text{если в первом испытании успех} \\ 0 & \text{иначе} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{если во втором испытании успех} \\ 0 & \text{иначе} \end{cases}$$

$Y = X_1 + X_2$ — общее кол-во успехов в серии. Найти закон распределения случайной величины Y.

X, Y	0	1
0	q^2	qp
1	pq	p^2

Тогда возможные значения случайной величины Y:

Y	0	1	2
P	q^2	$2pq$	p^2

2.4.2.2 Случай непрерывного случайного вектора

Если (X_1, X_2) — непрерывный случайный вектор, то функцию распределения случайной величины $Y = \varphi(X_1, X_2)$ можно найти по формуле

$$F_Y(y) = \iint\limits_{D(y)} f(x_1, x_2)\, dx_1\, dx_2$$

где f — совместная плотность распределения случайных величин X_1 и X_2, $D(y) = \{(x_1, x_2) \colon \varphi(x_1, x_2) < y\}$.

Доказательство.

$$
\begin{gathered}
F_Y(y) = P\{Y < y\} = \\
= \langle \text{ события } \{Y < y\} \text{ и } \{(X_1,\, X_2) \in D(y)\} \text{ эквивалентны } \rangle = \\
= \langle \text{ свойство непрерывного случайного вектора } \rangle = \\
= \iint\limits_{D(y)} f(x_1,\, x_2)\, dx_1\, dx_2
\end{gathered}
$$

2.4.3 Формула свёртки

Теорема. *Пусть*

1. $X_1,\, X_2$ — *непрерывные случайные величины;*
2. $X_1,\, X_2$ — *независимые случайные величины;*
3. $Y = X_1 + X_2$ *(т. е.* $\varphi(x_1,\, x_2) = x_1 + x_2$*).*

Тогда

$$
f_Y(y) = \int\limits_{-\infty}^{+\infty} f_{X_2}(y - x_1)\, f_{X_1}(x_1)\, dx_1
$$

Доказательство.

1.

$$
F_Y(y) = P\{Y < y\} = P\{X_1 + X_2 < y\} = \iint\limits_{D(y)} f(x_1,\, x_2)\, dx_1\, dx_2 =
$$

$$
= \left\langle X_1,\, X_2 \text{ — независимые } \implies f(x_1,\, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \right\rangle =
$$

$$
= \int\limits_{-\infty}^{+\infty} dx_1 \int\limits_{-\infty}^{y - x_1} f_{X_1}(x_1) f_{X_2}(x_2)\, dx_2 = \int\limits_{-\infty}^{+\infty} f_{X_1}(x_1) \left[F_{X_2}(x_2) \Big|_{-\infty}^{y - x_1} \right] dx_1 =
$$

$$
= \langle F_{X_2}(-\infty) = 0 \rangle = \int\limits_{-\infty}^{+\infty} f_{X_1}(x_1) F_{X_2}(y - x_1)\, dx_1
$$

2.

$$
f_Y(y) = \frac{d}{dy} F_Y(y) = \int\limits_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(y - x_1)\, dx_1
$$

Замечание 1. *Пусть f, $g\colon \mathbb{R} \to \mathbb{R}$ — абсолютно интегрируемые*[37] *функции.*

Определение. *Свёрткой функций f и g называется функция*

$$(f \circ g)(y) = \int\limits_{-\infty}^{+\infty} f(y - x) g(x)\, dx$$

Замечание 2. *Свёртка коммутативна, т. е. $(f \circ g)(y) = (g \circ f)(y)$*

Доказательство.

$$(f \circ g)(y) = \int\limits_{-\infty}^{+\infty} f(y - x) g(x) dx =$$

$$= \left\langle t = g - x;\ dx = -dt;\ x = +\infty \implies t = -\infty, x = -\infty \implies t = +\infty \right\rangle =$$

$$= -\int\limits_{+\infty}^{-\infty} f(t) g(t - y) dt = \int\limits_{-\infty}^{+\infty} f(t - y) f(t) dt = (g \circ f)(y)$$

Замечание 3. *С учётом этого обозначения, формула свёртки может быть записана в следующем виде:*

$$f_Y(y) = (f_{X_1} \circ f_{X_2})(y)$$

[37]Т. е. $\exists \int\limits_{-\infty}^{+\infty} |f(x)|\, dx < \infty$ для f, аналогично для g. — Прим. лект.

2.5 Числовые характеристики случайных величин

2.5.1 Математическое ожидание

Замечание. *Пусть на прямой задана система точек $x_1, \ldots, x_n$, массы которых $m_1, \ldots, m_n$ соответственно.*

$$x_{\textit{центра тяжести}} = \frac{\sum\limits_{i=1}^{n} m_i x_i}{\sum\limits_{i=1}^{n} m_i}$$

Замечание 2. *Если $f(x)$ — значение плотности бесконечного стержня в точке x, то*

$$x_{\textit{центра тяжести}} = \frac{\int\limits_{-\infty}^{+\infty} x\, f(x)\, dx}{\underbrace{\int\limits_{-\infty}^{+\infty} f(x)\, dx}_{\textit{масса стержня}}}$$

Определение. *Математическим ожиданием*[38] *дискретной величины X называют число*

$$M[X] = \sum_i p_i x_i$$

где i пробегает такое множество значений, что x_i исчерпывает все возможные значения случайной величины X; $p_i = P\{X = x_i\}$.

Замечание 1. *Если множество значений случайной величины X бесконечно (счётно), то указанный ряд должен сходиться абсолютно, т. е.*

$$\sum_i |x_i| \cdot p_i < \infty$$

В противном случае говорят, что $\nexists M[X]$.

Замечание 2. *Механический смысл математического ожидания.*

Будем интерпретировать величину p_i как «вероятностную массу» значения x_i случайной величины X. Т. к. $\sum\limits_i p_i = 1$, то $M[X]$ характеризует положение центра тяжести вероятностной массы.

[38] В английской литературе математическое ожидание называется Expected Value, обозначается как $E[X]$ или $E(X)$. — Прим. ред.

Определение. *Математическим ожиданием непрерывной случайной величины X называется число*[39]

$$M[X] = \int\limits_{-\infty}^{+\infty} x\, f(x)\, dx$$

где $f(x)$ — функция плотности случайной величины X.

Замечание 1. *В определении подразумевается, что указанный интеграл сходится абсолютно, т. е. его значение определено и конечно.*

Замечание 2. *Если интерпретировать функцию $f(x)$ как плотность материла стержня (из 1 кг вероятностной массы изготовлен стержень бесконечной длины), то MX — центр тяжести этого стержня.*

Пример 1. *Пусть* $X = \begin{cases} 1 & \text{с вероятностью } p \\ 0 & \text{с вероятностью } q = 1 - p \end{cases}$

$$\begin{array}{ccc} X & 0 & 1 \\ P & q & p \end{array}$$

Найдём MX: $MX = \sum\limits_{i=1}^{2} p_i x_i = 0 \cdot q + 1 \cdot p = p$

Пример 2. *X — непрерывная случайная величина. Предположим, что*

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

(X имеет распределение Коши).

$$MX = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty} \frac{x\,dx}{1+x^2} = \frac{1}{\pi}\int\limits_{-\infty}^{1} \frac{x\,dx}{1+x^2} + \frac{1}{\pi} \underbrace{\int\limits_{1}^{+\infty} \frac{x\,dx}{1+x^2}}_{\text{обозначим как интеграл } I}$$

Интеграл I расходится, т. к.

$$\frac{1}{1+x^2} \sim \frac{1}{x} \text{ при } x \to +\infty$$

а интеграл $\int\limits_{1}^{+\infty} \frac{dx}{x}$ *расходится.*

Таким образом, $\nexists MX$.

Лекция №15, 11.12.2018

[39] Если аргумент достаточно простой, то квадратные скобки в обозначении $M[X]$ часто опускают. Также часто MX обозначают просто буквой m. — Прим. ред.

2.5.1.1 Свойства математического ожидания

1. Если $P\{X = x_0\} = 1$ (т. е. если X фактически не является случайной), то $MX = x_0$;
2. $M[aX + b] = a \cdot MX + b$;
3. $M[X_1 + X_2] = MX_1 + MX_2$;
4. Если X_1, X_2 — независимы, то $M[X_1X_2] = (MX_1)(MX_2)$.

Замечание.

1. *Пусть X — случайная величина, а $\varphi\colon \mathbb{R} \to \mathbb{R}$ — некоторая функция, тогда*
   (a) *Если X — дискретная случайная величина, то*
   $$M[\varphi(X)] = \sum_i \varphi(x_i)\, p_i$$
   (b) *Если X — непрерывная случайная величина, то*
   $$M[\varphi(X)] = \int\limits_{-\infty}^{+\infty} \varphi(x)\, f(x)\, dx$$
2. *Если (X, Y) — случайный вектор, $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$, — функция, то*
   (a) *Если (X, Y) — дискретный случайный вектор, то*
   $$M[\varphi(X, Y)] = \sum_{i,j} \varphi(x_i, y_j)\, p_{ij}$$
   где $p_{ij} = P\{(X, Y) = (x_i, y_j)\}$.
   (b) *Если (X, Y) — непрерывный случайный вектор, то*
   $$M[\varphi(X, Y)] = \iint\limits_{R^2} \varphi(x, y)\, f(x, y)\, dx\, dy$$
   где f — совместная плотность распределения X и Y.

Доказательства.

1. *Ряд распределения:*

$$\begin{matrix} X & x_0 \\ P & 1 \end{matrix}$$

По определению: $MX = \sum\limits_i p_i x_i = 1 \cdot x_0 = x_0$.

2. *Докажем для случая непрерывной случайной величины.*

$$M[aX+b] = \langle \varphi(x) = ax+b;\ aX+b \text{ — непрерывная случайная величина } \rangle =$$
$$= \int\limits_{-\infty}^{+\infty} (ax+b)\, f(x)\, dx = a \int\limits_{-\infty}^{+\infty} x\, f(x)\, dx + b \int\limits_{-\infty}^{+\infty} f(x)\, dx = a \cdot MX + b.$$

3. *Доказательство для дискретного случая. Элементы X_1 обозначаются индексами i, пробегающими множество I; для X_2 используются j и J. Запись о:*

$$M[X_1+X_2] = \langle \varphi(x_1,\, x_2) = x_1 + x_2 \rangle =$$
$$= \sum_{i\in I}\sum_{j\in J} (x_{1,i} + x_{2,j}) p_{ij} = \sum_{i\in I}\sum_{j\in J} x_{1,i}\, p_{ij} + \sum_{i\in I}\sum_{j\in J} x_{2,j}\, p_{ij} =$$
$$= \sum_{i\in I} x_{1,i} \underbrace{\sum_{j\in J} p_{ij}}_{P\{X_1 = X_{1,i}\}} + \sum_{j\in J} x_{2,j} \underbrace{\sum_{i\in I} p_{ij}}_{P\{X_2 = x_{2,j}\}} = MX_1 + MX_2$$

4. *Докажем для непрерывных случайных величин*

$$M[X_1X_2] = \langle \varphi(x_1,\, x_2) = x_1\, x_2 \rangle = \iint\limits_{R^2} x_1\, x_2\, f(x_1,\, x_2)\, dx_1\, dx_2 =$$
$$= \langle X_1,\, X_2 \text{ — независимы } \implies f(x_1,\, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \rangle =$$
$$= \int\limits_{-\infty}^{+\infty} dx_1 \int\limits_{-\infty}^{+\infty} x_1\, x_2\, f_{X_1}(x_1)\, f_{X_2}(x_2)\, dx_2 =$$
$$\left(\int\limits_{-\infty}^{+\infty} x_1\, f_{X_1}(x_1)\, dx_1 \right) \left(\int\limits_{-\infty}^{+\infty} x_2\, f_{X_2}(x_2)\, dx_2 \right) = (MX_1)(MX_2)$$

Рубрика «Сделай сам». *Доказательства для дискретного случая свойств 2 и 4, непрерывного случая свойства 3 предлагается написать самостоятельно.*

Пусть $\overrightarrow{X} = (X_1,\, \ldots,\, X_n)$ — случайный вектор.

Определение. *Вектором математических ожиданий (вектором средних) случайного вектора называют*

$$\overrightarrow{MX} = (MX_1,\, \ldots,\, MX_n)$$

2.5.2 Дисперсия

Определение. *Дисперсией*[40] *случайной величины* X *называют число*

$$D[X] = M[(X-m)^2]$$

где $m = MX$.

Замечание 1. *Если* X *— дискретная случайная величина, то*

$$DX = \langle DX = M[(X-m)^2],\ \varphi(x) = (x-m)^2 \rangle = \sum_i (x_i - m)^2 p_i$$

где $p_i = P\{X = x_i\}$.

Замечание 2. *Если* X *— непрерывная случайная величина, то*

$$D[X] = \int\limits_{-\infty}^{+\infty} (x-m)^2 f(x)\, dx$$

где f *— функция плотности случайной величины* X.

Замечание 3. ***Механический смысл дисперсии.***

Дисперсия случайной величины характеризует разброс значений этой случайной величины относительно математического ожидания. Чем больше дисперсия, тем больше разброс значений.

С точки зрения механики дисперсия — момент инерции вероятностной массы относительно математического ожидания.

Пример. $X \sim B(1, p)$,

т. е. $X = \begin{cases} 1, & \text{с вероятностью } p \\ 0, & \text{с вероятностью } q = 1-p \end{cases}$

X	0	1
P	$1-p$	p

Математическое ожидание $MX = p$; *дисперсия:*

$$DX = \sum_i (x_i - m)^2 p_i = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p =$$
$$= p^2(1-p) + (1-p)^2 p = p(1-p)[p + 1 - p] = pq$$

[40]В английской литературе дисперсия называется Variance, обозначается как $Var[X]$ или $Var(X)$. — Прим. ред.

2.5.2.1 Свойства дисперсии

1. $DX \geqslant 0$;
2. Если $P\{X = x_0\} = 1$, то $DX = 0$.
3. $D[aX + b] = a^2 DX$;
4. $D[X] = M[X^2] - (MX)^2$;
5. Если X_1, X_2 — независимы, то $D[X_1 + X_2] = DX_1 + DX_2$.

Доказательства.

1. $DX = MY$, где $Y = (X - m)^2$. Т. к. $Y \geqslant 0$, то следует, что $DX = MY \geqslant 0$.

2.

$$\begin{matrix} X & x_0 \\ P & 1 \end{matrix}$$

Математическое ожидание $MX = m = x_0$.

Дисперсия $DX = \sum_i (x_i - m)^2 p_i = (x_0 - x_0)^2 \cdot 1 = 0$.

3.

$$D[aX + b] = M[[(aX + b) - M(aX + b)]^2] = M[[aX + b - a \cdot MX - b]^2] = \\ = M[(a(X - MX))^2] = a^2 M[(X - MX)^2] = a^2 DX$$

4. Обозначим $m = MX$; тогда

$$DX = M[(X - m)^2] = M[X^2 - 2mX + m^2] = M[X^2] - 2\underbrace{m \cdot M[X]}_{m^2} + m^2 = \\ = M[X^2] - m^2$$

5.

$$D(X_1 + X_2) = \langle \text{ по свойству 4 } \rangle = M[(X_1 + X_2)^2] - (M(X_1 + X_2))^2 = \\ = M[X_1^2] + M[X_2^2] + 2\,M[X_1 X_2] - (MX_1)^2 - (MX_2)^2 - 2\,MX_1 \cdot MX_2 = \\ = \langle X_1,\, X_2 \text{ — независимые, тогда } M[X_1\, X_2] = MX_1 \cdot MX_2 \rangle = \\ = (M[X_1^2] - (MX_1)^2) + (M[X_2^2] - (MX_2)^2) = DX_1 + DX_2$$

Замечание 1. *Можно доказать утверждение, обратное свойству 2: если дисперсия некоторой случайной величины равна 0, то X принимает единственное значение с вероятностью 1.*

Замечание 2. *Свойство 5 справедливо для любого числа **попарно** независимых случайных величин $X_1, \ldots, X_n$:*

$$D[X_1 + \ldots + X_n] = DX_1 + \ldots + DX_n$$

Замечание 3. *DX имеет размерность, равную квадрату размерности случайной величины X. Это не всегда удобно, особенно при решении практических задач. Поэтому рассматривают такую числовую характеристику, как среднеквадратичное отклонение (СКО).*

Определение. *Среднеквадратичным отклонением (СКО) случайной величины X называют число*

$$\sigma_X = \sqrt{D[X]}$$

2.5.3 Математические ожидания и дисперсия основных случайных величин

2.5.3.1 Биномиальная случайная величина

Обозначение $X \sim B(n, p)$, X — кол-во успехов в серии из n испытаний по схеме Бернулли с вероятностью успеха p. Рассмотрим случайную величину X_i, $i = \overline{1; n}$:

$$X_i = \begin{cases} 1, & \text{если в } i\text{-ом испытании произошёл успех,} \\ 0, & \text{если в } i\text{-ом имела место неудача} \end{cases}$$

Случайные величины $X_1, \ldots, X_n$ независимы, т. к. отдельные испытания в схеме Бернулли независимы.

Каждое $X_i \sim B(1, p)$, $i = \overline{1; n}$. Ранее было показано, что $MX_i = p$, $DX_i = pq$. Тогда

$$X = \sum_{i=1}^{n} X_i$$

$$MX = M\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} MX_i = np$$

$$DX = D\left[\sum_{i=1}^{n} X_i\right] = \langle X_i \text{ независимы}\rangle = \sum_{i=1}^{n} DX_i = npq$$

2.5.3.2 Пуассоновская случайная величина

$X \sim \Pi(\lambda)$

$P\{X = k\} = \dfrac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, \ldots$

Тогда математическое ожидание выражается

$$MX = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}\lambda}{(k-1)!} =$$

$$= \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \langle i = k-1\rangle = \lambda e^{-\lambda}\underbrace{\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}}_{\text{формула Маклорена для } e^{\lambda}} = \lambda$$

Дисперсия выражается как $DX = M[X^2] - (MX)^2$.
$MX^2 = \sum\limits_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \ldots = \lambda^2 + \lambda$. Тогда $DX = \lambda$.

2.5.3.3 Случайная величина X, имеющая геометрическое распределение с параметром p

$P\{X = k\} = pq^k,\ k = 0,\, 1,\, 2,\, \ldots$

$$MX = \sum_{k=0}^{\infty} k\cdot pq^k = \sum_{k=1}^{\infty} k\cdot pq^k = pq\underbrace{\sum_{k=1}^{\infty} kq^{k-1}}_{=1+2q+3q^2+4q^3+\ldots} =$$

$$=\Big\langle \text{ продифференцируем сумму бесконечно убывающей геометрической прогрессии:}$$

$$(1+q+q^2+\ldots)' = \left(\frac{1}{1-q}\right)' \Longrightarrow 1+2q+3q^2+4q^3+\ldots = \frac{1}{(1-q)^2} = \frac{1}{p^2}\Big\rangle =$$

$$= pq\frac{1}{p^2} = \frac{q}{p}$$

Аналогичным образом можно показать, что $DX = \dfrac{q}{p^2}$

2.5.3.4 Равномерно распределённая случайная величина

$X \sim R(a,\, b)$

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,\, b) \\ 0 & \text{иначе} \end{cases}$$

$$MX = \int\limits_{-\infty}^{\infty} x\, f(x)\, dx = \frac{1}{b-a}\int\limits_{a}^{b} x\, dx = \frac{1}{2(b-a)}(b^2 - a^2) = \frac{a+b}{2}$$

$$DX = \int\limits_{-\infty}^{+\infty} \left(x - \frac{a+b}{2}\right)^2 f(x)\,dx = \frac{1}{b-a}\int\limits_a^b \left(x - \frac{a+b}{2}\right)^2 dx =$$
$$= \frac{1}{3(b-a)}\left(x - \frac{a+b}{2}\right)^3\Bigg|_a^b = \frac{1}{3(b-a)}\left(\frac{(b-a)^3}{8} - \frac{(a-b)^3}{8}\right) =$$
$$= \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}$$

2.5.3.5 Экспоненциальная случайная величина

$X \sim Exp(\lambda)$

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

$$MX = \int\limits_{-\infty}^{+\infty} x\, f(x) x = \lambda \int\limits_0^{+\infty} xe^{-\lambda x} dx = -\int\limits_0^{+\infty} x de^{-\lambda x} = \cancelto{0}{-xe^{-\lambda x}\Big|_0^{+\infty}} + \int\limits_0^{+\infty} e^{-\lambda x} dx =$$
$$= -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^{+\infty} = \frac{1}{\lambda}$$

$DX = M[X^2] - (MX)^2,$

$$M[X^2] = \int\limits_{-\infty}^{+\infty} x^2\, f(x)\,dx = \lambda \int\limits_0^{+\infty} x^2 e^{-\lambda x} dx = -\int\limits_0^{+\infty} x^2 de^{-\lambda x} =$$
$$= \cancelto{0}{-x^2 e^{-\lambda x}\Big|_0^{+\infty}} + 2\underbrace{\int_0^{+\infty} xe^{-\lambda x} dx}_{=\frac{MX}{\lambda}} = \frac{2}{\lambda^2}$$

Таким образом,

$$DX = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

2.5.3.6 Нормальная случайная величина

$X \sim N(\underbrace{m}_{MX}, \underbrace{\sigma^2}_{DX})$

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

$$MX = \int\limits_{-\infty}^{+\infty} xf(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} xe^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = \langle x-m=t \rangle =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} (t+m)e^{-\frac{t^2}{2\sigma^2}}\,dt =$$

$$= \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} te^{-\frac{t^2}{2\sigma^2}}\,dt}_{0 \text{ (нечётная функция)}} + m \cdot \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma^2}}\,dt}_{1 \text{ (условие нормировки)}} = m$$

$$DX = M[(X-MX)^2] = \int\limits_{-\infty}^{+\infty} (x-m)^2 f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx =$$

$$= \left\langle \frac{x-m}{\sigma} = t,\, dx = \sigma dt \right\rangle =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}}\,dt = -\frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} t\,de^{-\frac{t^2}{2}} = \langle \text{ по частям } \rangle =$$

$$= \underbrace{-\frac{\sigma^2}{\sqrt{2\pi}} te^{-\frac{t^2}{2}} \Big|_{-\infty}^{+\infty}}_{0} + \sigma^2 \cdot \underbrace{\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\,dt}_{1 \text{ (усл. норм. } N(0,\,1))} = \sigma^2$$

Лекция №16, 18.12.2018

2.5.4 Моменты

Пусть X — случайная величина.

Определение. *Моментом k-ого порядка (k-ым начальным моментом) случайной величины X называется число*

$$m_k = M[X^k]$$

Определение. *Центральным моментом k-ого порядка случайной величины X называется число*

$$\mathring{m}_k = M[(X-m)^k]$$

где $m = MX$.

Замечание 1. *Если X — дискретная случайная величина, то*

$$m_k = \sum_i x_i^k p_i$$

$$\mathring{m}_k = \sum_i (x_i - m)^k p_i$$

Если X — непрерывная случайная величина, то

$$m_k = \int\limits_{-\infty}^{+\infty} x^k f(x)\,dx$$

$$\mathring{m}_k = \int\limits_{-\infty}^{+\infty} (x - m)^k f(x)\,dx$$

где f — функция плотности случайной величины X.

Замечание 2. $m_0 = MX$*;* $\mathring{m}_2 = DX$*;* $\mathring{m}_1 = M[(X-m)^1] = MX - m = 0$*.*

2.5.5 Кванти́ль

Пусть

1. X — случайная величина;
2. $\alpha \in (0,\,1)$.

Определение. *Кванти́лью уровня α (α-квантилью) случайной величины X называется число q_α такое, что*

$$P\{X < q_\alpha\} \leqslant \alpha,\ P\{X > q_\alpha\} \leqslant 1$$

Замечание. *Для непрерывной случайной величины X квантиль уровня α является решением уравнения*

$$F(x) = \alpha$$

где F — функция распределения случайной величины X.

Замечание. *Если f — функция плотности случайной величины X, то слева от точки q_α «находится» α кг массы вероятности, а справа $(1-\alpha)$ кг.*

Пример. $X \sim Exp(\lambda)$*, найдём квантиль уровня α случайной величины X.*

Решение.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

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

У нас уравнение

$$1 - e^{-\lambda x} = \alpha \implies e^{-\lambda x} = 1 - \alpha \implies -\lambda x = \ln(1-\alpha) \implies x = -\frac{\ln(1-\alpha)}{\lambda}$$

Ответ. $q_\alpha = -\frac{\ln(1-\alpha)}{\lambda}$.

Определение. *Медианой случайной величины X называют её квантиль уровня $\frac{1}{2}$.*

Пример. *Медиана случайной величины $X \sim Exp(\lambda)$ (см. пример выше)*

$$q_{\frac{1}{2}} = -\frac{\ln(1-\frac{1}{2})}{\lambda} = \frac{\ln 2}{\lambda}$$

2.5.6 Ковариации

До сих пор мы рассматривали числовые характеристики одномерных случайных величин. Ковариация является характеристикой случайного вектора.

Пусть (X, Y) — двумерный случайный вектор.

Определение. *Ковариацией случайных величин X и Y называется число*

$$\text{cov}(X, Y) = M[(X - m_X)(Y - m_Y)]$$

где $m_X = MX$, $m_Y = MY$.

Замечание. *Из определения следует:*

1. *В случае дискретного случайного вектора*

$$\text{cov}(X, Y) = \sum_i \sum_j (x_i - m_X)(y_j - m_Y)p_{ij}$$

где $p_{ij} = P\{(X, Y) = (x_i, y_j)\}$.

2. *В случае непрерывного случайного вектора*

$$\text{cov}(X, Y) = \iint\limits_{R^2} (x - m_X)(y - m_Y) f(x, y)\, dx\, dy$$

где f — совместная плотность величин X и Y.

2.5.6.1 Свойства ковариации

1. $D(X+Y) = DX + DY + 2\operatorname{cov}(X, Y)$;
2. $\operatorname{cov}(X, X) = DX$;
3. Если X, Y — независимые, то $\operatorname{cov}(X, Y) = 0$;
4. $\operatorname{cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\operatorname{cov}(X, Y)$
5. $|\operatorname{cov}(X, Y)| \leqslant \sqrt{DX \cdot DY}$, причём $|\operatorname{cov}(X, Y)| = \sqrt{DX \cdot DY} \iff \exists\exists a, b \in \mathbb{R}$, $Y = aX + b$ (т. е. X и Y связаны линейной зависимостью);
6. $\operatorname{cov}(X, Y) = M[XY] - (MX)(MY)$.

Доказательства.

1.

$$\begin{gathered}D(X+Y) = M[((X+Y) - M[X+Y])^2] = \langle MX = m_1,\, MY = m_2\rangle = \\ = M[((X-m_1) - M[Y-m_2])^2] = \\ = \underbrace{M[(X-m_1)^2]}_{=DX} + \underbrace{M[(Y-m_2)^2]}_{=DY} + 2\underbrace{M[(X-m_1)(Y-m_2)]}_{\operatorname{cov}(X,Y)} = \\ = DX + DY + 2\operatorname{cov}(X, Y)\end{gathered}$$

2. $\operatorname{cov}(X, X) = M[(X-m)(X-m)] = M[(X-m)^2] = DX$.

3.

$$\begin{gathered}\operatorname{cov}(X, Y) = M[(X-m_1)(Y-m_2)] = \\ = \langle X, Y \text{ — \textit{независимы}} \implies (X-m_1) \text{ \textit{и} } (Y-m_2) \text{ \textit{тоже независимы}}\rangle = \\ = [\cancelto{0}{M(X-m_1)}][\cancelto{0}{M(Y-m_2)}] = 0\end{gathered}$$

4.

$$\begin{gathered}\operatorname{cov}(a_1X + b_1, a_2X + b_2) = \\ = M[[a_1X + b_1 - M(a_1X + b_1)] \cdot [a_2X + b_2 - M(a_2X + b_2)]] = \\ = M[[a_1X + \cancel{b_1} - a_1m_1 - \cancel{b_1}][a_2X + \cancel{b_2} - a_2m_2 - \cancel{b_2}]] = \\ = M[a_1a_2(X_1 - MX_1)(X_2 - MX_2)]\end{gathered}$$

5. (a) *Выберем произвольное число $t \in R$. Рассмотрим случайную величину $Z(t) = tX - Y$.*

Тогда $D[Z(t)] = D[tX - Y] = \langle$ *свойство 1 + свойство дисперсии* $\rangle = t^2 DX + DY - 2t\operatorname{cov}(X, Y) = DX \cdot t^2 - 2t \cdot \operatorname{cov}(X, Y) + DY$ *— квадратный трёхчлен относительно* t.

Т. к. $D[Z(t)] \geqslant 0$, *следовательно, трёхчлен должен быть параболой вверх, следовательно — дискриминант* $D \leqslant 0$.

$$\frac{D}{4} = (\operatorname{cov}(X, Y))^2 - DX \cdot DY \leqslant 0 \implies |\operatorname{cov}(X, Y)| \leqslant \sqrt{DX \cdot DY}$$

(b) ***Необходимость*** $(\implies)$.

Если

$$|\operatorname{cov}(X, Y)| = \sqrt{DX \cdot DY} \implies \text{дискриминант} = 0 \implies$$
$$\implies D[Z(t)] \text{ имеет единственный корень. Обозначим его } t = a \implies$$
$$\implies D[Z(a)] = 0 \implies$$
$$\implies Z(a) = aX - Y \text{ принимает единственное значение с вероятностью 1,}$$
$$\text{обозначим это значение как} - b \implies Z(a) = aX - Y = -b \implies$$
$$\implies Y = aX + b$$

(c) ***Достаточность*** $(\impliedby)$.

Если $Y = aX+b \implies Z(a) = -b \implies D[Z(a)] = 0 \implies$ *дискриминант* $= 0 \implies |\operatorname{cov}(X, Y)| = \sqrt{DX \cdot DY}$.

6.

$$\operatorname{cov}(X, Y) = M[(X - m_1)(Y - m_2)] = M[XY - m_1Y - m_2X + m_1m_2] =$$
$$= M[XY] - m_1 \underbrace{MY}_{m_2} - m_2 \underbrace{MX}_{m_1} + m_1m_2 = M[XY] - m_1m_2$$

Замечание 1. *Свойство 1 с учётом 4 допускает обобщение:*

$$D(a_1X + a_2Y + b) = a_1^2 DX + a_2^2 DY + 2a_1a_2\operatorname{cov}(X, Y)$$

Замечание 2. *Пусть* $Y = aX + b$. *В соответствии со свойством 4* $\operatorname{cov}(X, Y) = \operatorname{cov}(X, aX + b) = aDX$, *следовательно, знак коэффициента* a *совпадает со знаком* $\operatorname{cov}(X, Y)$.

Поэтому свойство 5 можно уточнить

$$|\operatorname{cov}(X, Y)| = \sqrt{DX \cdot DY} \iff Y = aX + b$$

где $a > 0$, *если* $\operatorname{cov}(X, Y) > 0$; $a < 0$, *если* $\operatorname{cov}(X, Y) < 0$.

Или: $\operatorname{cov}(X, Y) = \begin{cases} \sqrt{DX \cdot DY}, & \text{если } a > 0 \\ -\sqrt{DX \cdot DY}, & \text{если } a < 0 \end{cases}$

Определение. *Случайные величины X и Y называют некоррелированными, если* $\operatorname{cov}(X, Y) = 0$.

Замечание 1. *Из свойства 3 следует, что если X, Y — независимые, то X, Y — некоррелированные. Обратное неверно — приведите пример самостоятельно.*

Замечание 2. *Недостатком ковариации является то, что она имеет размерность, равную произведению разностей случайных величин X и Y. Часто рассматривают аналогичную безразмерную характеристику, которая называется коэффициентом корреляции случайных величин случайных величин X и Y:*

$$\rho(X, Y) = \rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{DX \cdot DY}}$$

где $DX \cdot DY > 0$.

2.5.6.2 Свойства коэффициента корреляции

1. $\rho_{XX} = 1$;
2. Если X, Y независимые, то $\rho_{XY} = 0$;
3. $\rho(a_1 X + b_1,\, a_2 Y + b_2) = \pm\rho(X, Y)$, причём $\pm$ заменяется на
   - $+$, если $a_1 a_2 > 0$;
   - $-$, если $a_1 a_2 < 0$.
4. $|\rho_{XY}| \leqslant 1$, причём $\rho_{XY} = \begin{cases} 1, & \text{когда } Y = aX + b, \text{ где } a > 0, \\ -1, & \text{когда } Y = aX + b, \text{ где } a < 0 \end{cases}$

Доказательство. *Все свойства следуют из свойств ковариации (докажите самостоятельно).*

Замечание. *Из свойств коэффициента корреляции следует, что*

1. *Если X, Y — независимы, то $\rho_{XY} = 0$;*
2. $\rho_{XY} = 1 \iff Y = aX + b$.

Таким образом, коэффициент корреляции выражает степень линейной зависимости случайных величин X и Y. $\rho \lesssim 1$; чем ближе к единице значение ρ, тем взаимосвязь X и Y больше похожа на прямую.

Пусть $\overrightarrow{X} = (X_1, \ldots, X_n)$ — случайный вектор.

Определение. *Ковариантной матрицей случайного вектора $\overrightarrow{X}$ называется матрица*

$$\Sigma_{\overrightarrow{X}} = (\sigma_{ij})_{i,\, j = \overline{1;\, n}}$$

где $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$.

Замечание. *Логично показать некоторые свойства ковариантной матрицы:*

1. $\sigma_{ii} = DX_i$;

2. $\Sigma_{\overrightarrow{X}} = \Sigma_{\overrightarrow{X}}^T$;

3. *Если*
$$\overrightarrow{Y} = \overrightarrow{X}B + \overrightarrow{c}$$
где $\overrightarrow{Y} = (Y_1, \ldots, Y_m)$, $\overrightarrow{X} = (X_1, \ldots, X_n)$, $B \in M_{n,m}(\mathbb{R})$ *(т. е.* $\overrightarrow{Y}$ *является линейной функцией от вектора* $\overrightarrow{X}$*), то*
$$\Sigma_{\overrightarrow{Y}} = B^T \Sigma_{\overrightarrow{X}} B$$

4. *Матрица* $\Sigma_{\overrightarrow{X}}$ *является неотрицательно определённой, т. е.* $\forall\, \overrightarrow{b} \in \mathbb{R}^{\omega}$
$$\overrightarrow{b}^T \Sigma_{\overrightarrow{X}} \overrightarrow{b} \geqslant 0$$

5. *Если все компоненты вектора* $\overrightarrow{X}$ *попарно независимы, то* $\Sigma_{\overrightarrow{X}}$ *— диагональная матрица.*

Доказательство. *Без доказательства.*

Определение. *Корреляционной матрицей вектора* $\overrightarrow{X} = (X_1, \ldots, X_n)$ *называют матрицу*
$$P = (\rho_{ij})_{i,\, j = \overline{1;\, n}}$$
где $\rho_{ij} = \rho(X_i,\, X_j)$.

2.5.7 Условные числовые характеристики

Пусть $(X,\, Y)$ — двумерный случайный вектор. Рассмотрим распределение компоненты X этого вектора при условии, что $Y = y$. Т. к. условное распределение обладает всеми свойствами безусловного распределения, то для него также можно рассмотреть числовые характеристики.

Дискретный случай. $(X,\, Y)$ — дискретный случайный вектор. Ранее мы находили, что условное распределение компоненты X при условии $Y = y_j$
$$\pi_{ij} = P\{X = x_i \,|\, Y = y_j\} = \frac{p_{ij}}{P_{Y_j}}$$

Определение. *Значением условного математического ожидания случайной величины* X *при условии* $Y = y_j$ *называется число*
$$M[X \,|\, Y = y_j] = \sum_i \pi_{ij} x_i$$

Замечание. *Значение условного математического ожидания $M[Y \mid X = x_i]$ определяется аналогично.*

Непрерывный случай. Пусть (X, Y) — непрерывный случайный вектор. Ранее мы определили условную плотность

$$f_X(x \mid Y = y) = \frac{f(x, y)}{f_Y(y)}$$

Определение. ***Значением*** *условного математического ожидания случайной величины X при условии $Y = y$ называют число*

$$M[X \mid Y = y] = \int\limits_{-\infty}^{+\infty} x f_X(x \mid Y = y)\, dx$$

Определение. *Условным математическим ожиданием случайной величины относительно случайной величины Y называют функцию*

$$g(Y) = M[X \mid Y]$$

которая

1. *Имеет область определения, совпадающую со множеством значений случайной величины Y;*

2. *Для каждого возможного значения y случайной величины Y значение $g(Y) = M[X \mid Y = y]$ является значением условного математического ожидания.*

Замечание 1. *Условное математическое ожидание является функцией случайной величины Y, т. е. оно само является случайной величиной.*

Замечание 2. *Условное математическое ожидание $M[Y \mid X]$ определяется аналогично.*

Определение. *Условной дисперсией случайной величины X относительно случайной величины Y называют случайную величину*

$$D[X \mid Y] = M[(X - M[X \mid Y])^2]$$

Замечание. *Значение условной дисперсии*

- *Для дискретного случайного вектора*

$$D[X \mid Y = y_j] = \sum_i \pi_{ij}(x_i - M[X \mid Y = y_j])^2$$

- *Для непрерывного случайного вектора*

$$D[X\,|\,Y=y_j]=\int\limits_{-\infty}^{+\infty}(x-M[X\,|\,Y=y_j])^2 f_X(x\,|\,Y=y_j)\,dx$$

Весенний семестр
Лекция №17, 11.02.2019

2.6 Предельные теоремы теории вероятностей

2.6.1 Неравенства Чебышева

Теорема. ***Первое неравенство Чебышева.***
Пусть

1. X — *случайная величина;*
2. $X \geqslant 0$ *(т. е.* $P\{X<0\}=0$*);*
3. $\exists MX$.

Тогда

$$\forall\varepsilon>0\;\; P\{X\geqslant\varepsilon\}\leqslant\frac{MX}{\varepsilon}$$

Доказательство. *(для случая непрерывной случайной величины* X*, для дискретной случайной величины* X *доказательство аналогично).*

$$MX=\int\limits_{-\infty}^{+\infty}x\,f(x)\,dx=\langle\textit{свойство аддитивности определённого интеграла}\rangle=$$

$$=\int\limits_{0}^{\varepsilon}x\,f(x)\,dx+\int\limits_{\varepsilon}^{+\infty}x\,f(x)\,dx\geqslant\int\limits_{\varepsilon}^{+\infty}x\,f(x)\,dx\geqslant\langle\forall x\in[\varepsilon;\,\infty)\;x\geqslant\varepsilon\rangle\geqslant$$

$$\geqslant\varepsilon\int\limits_{\varepsilon}^{+\infty}f(x)\,dx=\varepsilon\,P\{X\geqslant\varepsilon\}$$

Теорема. ***Второе неравенство Чебышева.***
Пусть

1. X — *случайная величина;*

2. $\exists MX$, $\exists DX$.

Тогда

$$\forall\varepsilon > 0 \;\; P\{|X - MX| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}$$

Доказательство.

$$DX = M[(X - MX)^2] \geqslant$$

$\langle$*Рассмотрим случайную величину* $Y = (X - MX)^2 \geqslant 0$*; из первого неравенства Чебышева для* Y *следует, что* $\forall\sigma > 0$ $P\{Y \geqslant \sigma\} \leqslant \dfrac{MY}{\sigma}$. *Используем это для* $\sigma = \varepsilon^2\rangle$

$$\geqslant \sigma P\{(X - MX)^2 \geqslant \sigma\} = \langle\sigma = \varepsilon^2\rangle = \varepsilon^2 P\{(X - MX)^2 \geqslant \varepsilon^2\} = \varepsilon^2 P\{|X - MX| \geqslant \varepsilon\}$$

Таким образом,

$$DX \geqslant \varepsilon^2 P\{|X - MX| \geqslant \varepsilon\} \implies P\{|X - MX| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}$$

Задача. *Значение предельно допустимого давления в пневмокамере ракеты составляет* 200 *Па. После проверки большого кол-ва ракет было установлено, что среднее значение давления в пневмокамерах составляет* 150 *Па. Оценить вероятность того, что давление в очередной проверенной камере ракеты превысит предельно допустимое значение, если также известно, что среднеквадратичное отклонение значения давления составляет* 5 *Па.*

Решение.

- *Пусть* X *— случайная величина, принимающая значения, равные значениям давления в пневмокамерах этих ракет (в Па). Тогда* $MX = 150$, $DX = 5^2$.

- *Используем для оценки первое неравенство Чебышева, т. к.* $X \geqslant 0$:

$$P\{X \geqslant 200\} \leqslant \frac{MX}{200} = \frac{3}{4}$$

- *Используем второе неравенство Чебышева:*

$$P\{X \geqslant 200\} = P\{X - MX \geqslant 200 - 150\} = P\{X - MX \geqslant 50\} \leqslant$$
$$\leqslant P\left\{\{X - MX \geqslant 50\} + \{X - MX \leqslant -50\}\right\} = P\{|X - MX| \geqslant \underbrace{50}_{\varepsilon}\} \leqslant$$
$$\leqslant \langle\textit{второе неравенство Чебышева}\rangle \leqslant \frac{DX}{50^2} = \frac{5^2}{50^2} = \frac{1}{100}$$

Замечание 1. *В результате решения примера получено:*

- *С использованием первого неравенства Чебышева:*

$$P\{X \geqslant 200\} \leqslant \frac{3}{4}$$

- *С использованием второго неравенства Чебышева:*

$$P\{X \geqslant 200\} \leqslant \frac{1}{100}$$

Второе неравенство Чебышева даёт существенно более точный результат, т. к. дополнительно использует информацию о значении дисперсии.

Замечание 2. *Очевидно, что*

1. *Если в первом неравенстве Чебышева $\varepsilon \leqslant MX$, то оно даёт тривиальную оценку: $P \leqslant 1$;*

2. *Если во втором неравенстве Чебышева $\varepsilon \leqslant \sqrt{DX}$, то оно также даёт тривиальную оценку.*

2.6.2 Виды сходимости последовательности случайных величин

Пусть X_1, X_2, … — последовательность случайных величин. Мы изучим лишь два вида сходимости:

1. Сходимость по вероятности;

2. Слабая сходимость.

Определение. *Говорят, что последовательность случайных величин X_1, X_2, … сходится по вероятности к случайной величине Z, если*

$$\forall \varepsilon > 0 \; P\{|X_n - Z| \geqslant \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

Обозначается как $X_n \xrightarrow[n\to\infty]{P} Z$.

Определение. *Говорят, что последовательность величин X_1, X_2, … слабо сходится к случайной величине Z, если функциональная последовательность*

$$F_{X_1}(x),\; F_{X_2}(x),\; \ldots$$

поточечно сходится к функции $F_Z(x)$ во всех точках непрерывности последней, т. е.

$$(\forall x_0 \in \mathbb{R})((\text{функция } F_Z(x) \text{ непрерывна в } x_0) \implies \underbrace{F_{X_n}(x_0)}_{\text{числовая последовательность}} \xrightarrow[n\to\infty]{} F_Z(x)$$

Обозначается как $F_{X_n}(x) \implies F_Z(x)$.

2.6.3 Закон больших чисел

Определение. *Говорят, что последовательность X_1, X_2, ... удовлетворяет закону больших чисел, если*

$$\forall \varepsilon > 0 \; P\Big\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} m_i \right| \geqslant \varepsilon \Big\} \xrightarrow[n \to \infty]{} 0$$

где $m_i = MX_i$, $i \in \mathbb{N}$.

Замечание 1. *Рассмотрим случайную величину $\overline{X_n} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$.*

$$M\overline{X_n} = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} MX_i = \frac{1}{n}\sum_{i=1}^{n} m_i$$

Выполнение закона больших чисел для последовательности X_1, X_2, ... означает, что

$$\forall \varepsilon > 0 \; P\{|\overline{X_n} - M\overline{X_n}| \geqslant \varepsilon\} \xrightarrow[n \to \infty]{} P$$

т. е. при достаточно больших n случайная величина $\overline{X_n}$ принимает детерминированное (т. е. неслучайное) значение $\frac{1}{n} \sum\limits_{i=1}^{n} m_i$ с вероятностью, близкой к 1. Таким образом, в этом случае случайная величина $\overline{X_n}$ при достаточно больших n утрачивает случайный характер.

Замечание 2. *Выполнение закона больших чисел для последовательности X_1, X_2, ... означает, что последовательность*

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_i)$$

при $n \to \infty$ сходится по вероятности к нулю, т. е.

$$\forall \varepsilon > 0 \; P\{|Y_n - 0| \geqslant \varepsilon\} \xrightarrow[n \to \infty]{} 0$$

Вопрос: какой должна быть последовательность X_1, X_2, ..., чтобы она удовлетворяла закону больших чисел?

Теорема. *(Чебышева, достаточное условие того, что последовательность удовлетворяет закону больших чисел)*

Пусть

1. *X_1, X_2, ... — последовательность независимых случайных величин;*
2. *$\exists MX_i = m_i$, $\exists DX_i = \sigma_i^2$, $i \in \mathbb{N}$;*

3. *Дисперсии случайных величин* X_1, X_2, … *ограничены в совокупности, т. е.* $\exists C > 0\ \sigma_i^2 \leqslant C,\ i \in \mathbb{N}$.

Тогда последовательность X_1, X_2, … *удовлетворяет закону больших чисел.*

Доказательство.

- *Рассмотрим* $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, $n \in \mathbb{N}$. *Тогда*

$$M\left[\overline{X_n}\right] = \frac{1}{n}\sum_{i=1}^{n} m_i$$

$$D\left[\overline{X_n}\right] = D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} D\left[\sum_{i=1}^{n} X_i\right] = \langle X_i \text{ независимы}\rangle =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$$

- *Применим к случайной величине* $\overline{X_n}$ *второе неравенство Чебышева:*

$$P\left\{|\overline{X_n} - M\overline{X_n}| \geqslant \varepsilon\right\} \leqslant \frac{D\overline{X_n}}{\varepsilon^2}$$

Таким образом,

$$P\Big\{ \left|\overline{X_n} - \frac{1}{n}\sum_{i=1}^{n} m_i\right| \geqslant \varepsilon\Big\} \leqslant \frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \tag{13}$$

$$\sum_{i=1}^{n}\sigma_i^2 \leqslant \langle \sigma_i^2 \leqslant C\rangle \leqslant \sum_{i=1}^{n} C = nC$$

$$0 \leqslant P\Big\{ \left|\overline{X_n} - \frac{1}{n}\sum_{i=1}^{n} m_i\right| \geqslant \varepsilon\Big\} \leqslant \frac{C}{\varepsilon^2}\cdot\frac{1}{n^2}\cdot n = \overset{0,\ n\to\infty}{\frac{C}{\varepsilon^2}\cdot\frac{1}{n}}$$

При $n \to \infty$ *по теореме о двух миллиционерах*

$$P\Big\{ \left|\overline{X_n} - \frac{1}{n}\sum_{i=1}^{n} m_i\right| \geqslant \varepsilon\Big\} \xrightarrow[n\to\infty]{} 0$$

т. е. последовательность X_1, X_2, … *удовлетворяет закону больших чисел.*

Замечание 1. *Если последовательность* X_1, X_2, … *удовлетворяет условию теоремы Чебышева, то, разумеется, она удовлетворяет закон больших чисел. При этом говорят, что эта последовательность удовлетворяет закону больших чисел в форме Чебышева.*

Замечание 2. *Условие 3 теоремы Чебышева можно ослабить, заменив его на условие 3′:*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} DX_i = 0$$

Из оценки 13 следует, что в этом случае последовательность X_1, X_2, … также удовлетворяет закону больших чисел.

Следствие 1. *Пусть*

1. *Выполнены условия теоремы Чебышева;*
2. *Все случайные величины X_i одинаково распределены (обозначим $m \equiv m_i \equiv MX_i$).*

   Тогда

$$\forall \varepsilon > 0 \; P\Big\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - m \right| \geqslant \varepsilon \Big\} \xrightarrow[n\to\infty]{} 0$$

Доказательство. *Т. к. $m_i \equiv m$, то $\frac{1}{n}\sum\limits_{i=1}^{n} m_i = m$ и используем закон больших чисел в форме Чебышева.*

Следствие 2. ***Закон больших чисел в форме Бернулли.***
Пусть

1. *Проводится n испытаний по схеме Бернулли с вероятностью успеха p;*
2. $$r_n = \frac{\{\textit{число «успехов» в этой серии}\}}{n}$$

   (r_n называют относительной (наблюдательной) частотой успеха)

Тогда $r_n \xrightarrow[n\to\infty]{P} P$.

Доказательство.

- *Введём случайные величины X_i, $i = \overline{1;\, n}$,*

$$X_i = \begin{cases} 1, & \textit{если в } i\textit{-ом испытании прозошёл успех} \\ 0 & \textit{иначе} \end{cases}$$

Тогда

1. *Закон распределения случайной величины X_i:*

$$\begin{array}{ccc} X_i & 0 & 1 \\ P & q = 1 - p & p \end{array}$$

Таким образом, все X_i одинаково распределены;

2. *$DX_i \equiv pq$ — ограничены в совокупности;*
3. *X_i независимы, т. к. отдельные испытания в схеме Бернулли независимы;*
4. *Таким образом, последовательность X_1, X_2, ... удовлетворяет первому следствию из теоремы Чебышева и для неё справедливо*

$$\forall\varepsilon > 0\ P\Bigg\{\Bigg|\frac{1}{n}\underbrace{\sum_{i=1}^{n} X_i}_{=r_n,\ \text{общее число успехов из } n \text{ испытаний}} - \underbrace{m}_{=P}\Bigg| \geqslant \varepsilon\Bigg\} \xrightarrow[n\to\infty]{} 0$$

$$\forall\varepsilon > 0\ P\{|r_n - P| \geqslant \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

т. е. $r_n \xrightarrow[n\to\infty]{P} P$.

Лекция №18, 18.02.2019

2.6.4 Центральная предельная теорема

Пусть:

1. X_1, X_2, ... — последовательность независимых случайных величин;
2. Все случайные величины из этой последовательности одинаково распределены;
3. $\exists MX_i = m$, $\exists DX_i = \sigma^2$, $i \in \mathbb{N}$.

Рассмотрим случайную величину $\overline{X_n} = \frac{1}{n}\sum\limits_i X_i$, т. е. $\overline{X}_1 = X_1$, $\overline{X}_2 = \frac{1}{2}(X_1 + X_2)$ и т. д. Тогда $M\overline{X_n} = m,\ n \in \mathbb{N}$; $D\overline{X_n} = \frac{\sigma^2}{n}, n \in \mathbb{N}$. Рассмотрим случайную величину

$$Y_n = \frac{\overline{X_n} - M\overline{X_n}}{\sqrt{D\overline{X_n}}} = \frac{\overline{X_n} - m}{\left(\dfrac{\sigma}{\sqrt{n}}\right)},\ n \in \mathbb{N}$$

Для неё $MY_n = 0$, $DY_n = 1$.

Теорема. ***Центральная предельная теорема.***

Пусть выполнены условия 1-3.

Тогда последовательность величин Y_n при $n \to \infty$ слабо сходятся к случайной величине Z, имеющей стандартное нормальное распределение, т. е.

$$\forall x \in \mathbb{R}\ F_Y(x) \xrightarrow[n\to\infty]{} F_Z(x)$$

где $Z \sim N(0,1)$, т. е. $F_Z(x) = \Phi(x) = \dfrac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{\frac{-t^2}{2}}\,dt$.

Доказательство. *Без доказательства.*

Замечание. *В этом случае говорят, что случайные величины Y_n имеют асимптотически стандартное нормальное распределение.*

Пример. *ЭВМ проводят суммирование 10^4 чисел, каждое из которых округлено до 10^{-4}. Считая, что ошибки округления отдельных слагаемых независимы и имеют равномерное распределение на отрезке $[-0.5 \cdot 10^4;\ 0.5 \cdot 10^{-4}]$ найти диапазон, в котором с вероятностью в 0.95 будет заключена ошибка результата суммирования.*

Решение.

1. *Пусть X_i — ошибка округления i-ого слагаемого, $i = \overline{1;\ 10^4}$;* $X_i \sim R[-0.5 \cdot 10^{-4};\ 0.5 \cdot 10^{-4}] \implies MX_i = 0; DX_i = (0.5 \cdot 10^{-4} - (-0.5 \cdot 10^{-4}))^2 \cdot \frac{1}{12} = \underbrace{\frac{1}{12}}_{\approx \frac{1}{10}} 10^{-8} \approx 10^{-9}$

2. *Тогда погрешность суммы всех таких слагаемых $S = \sum\limits_i X_i$, где $n = 10^4$.*

3. *Рассмотрим случайную величину $Y_n = \frac{\overline{X_n} - MX_i}{\sqrt{D\overline{X_n}/n}} = \frac{\overline{X_n}}{\sqrt{10^{-9}/10^4}} = \frac{\overline{X_n}}{\sqrt{10^{-13}}}$, $n = 10^4$.*

   Т. к. $n \gg 1$, то в соответствии с ЦПТ можно полагать, что

   $$Y_n \underbrace{\sim}_{\text{(приближённо)}} N(0,\ 1)$$

4. *$u_{0.975}$ — квантиль уровня 0.975 стандартного нормального распределения, из таблицы $u_{0.975} = 1.96$ Таким образом, $P\{|Y_n| < u_0.975\} = 0.95$. Таким образом, с вероятностью 0.95 можно утверждать, что*
   $\left|\frac{\overline{X_n}}{\sqrt{10^{-13}}}\right| < 1.96 \iff \left|\frac{1}{n}\overline{X_n}\right| < 1.96 \cdot \sqrt{10^{-13}} \iff \left|\overline{X_n} = \frac{1}{n}S\right| \iff |\sigma| \leqslant 1.96 \cdot n \cdot \sqrt{10^{-13}} \approx 6.2 \cdot 10^{-3}$

Ответ. *С вероятностью 0.95 можно утверждать, что погрешность результата не превышает $6.2 \cdot 10^{-3}$.*

Теорема. *Центральная теорема Муавра-Лапласа.*
Пусть

1. *Проводится большое число испытаний n по схеме Бернулли с вероятностью успеха p;*

2. *k — число успехов в этой серии.*

Тогда

$$P\{k_1 \leqslant k \leqslant k_2\} = \Phi(x_2) - \Phi(x_1)$$

где $x_i = \frac{k_i - np}{\sqrt{npq}}$, $i = \overline{1;\, 2}$, $q = 1 - p$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$.

Доказательство.

1. *Пусть X_i — случайная величина, принимающая значения 0 или 1 в соответствии с правилом*

$$X_i = \begin{cases} 1, & \textit{если в } i\textit{-ом испытании был успех;} \\ 0 & \textit{иначе} \end{cases}, \ i = \overline{1;\, n}$$

Тогда

   (a) *Случайные величины $X_1, \ldots, X_n$ независимы;*

   (b) $MX_i = p$, $DX_i = pq$, $i \in \mathbb{N}$;

   (c) *X_i одинаково распределены.*

2.

$$P\{k_1 \leqslant k \leqslant k_2\} = P\{k_1 \leqslant \sum_{i=1}^{n} X_i \leqslant k_2\} = P\left\{\frac{k_1}{n} - p \leqslant \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{=\overline{X_n}} - p \leqslant \frac{k_2}{n} - p\right\} =$$

$$= P\{\underbrace{\frac{\frac{k_1}{n} - p}{\sqrt{pqn}}}_{=X_1} \leqslant \underbrace{\frac{\overline{X_n} - \overbrace{p}^{=MX_i}}{\underbrace{\sqrt{pqn}}_{=DX_i}}}_{=Y_n} \leqslant \underbrace{\frac{\frac{k_2}{n} - p}{\sqrt{pqn}}}_{=X_2}\} =$$

$$= \langle \textit{в соответствии с ЦПТ (т. к. } n \gg 1) \ Y_n \sim N(0,\, 1) \textit{ приближённо} \rangle =$$
$$= \Phi(x_2) - \Phi(x_1)$$

Задача. *В эксперименте Пирсона по подбрасыванию симметричной монеты* 24000 *раз герб выпал* 12012 *раз. Какова вероятность того, что при повторном испытании (повторении эксперимента) абсолютное отклонение относительной частоты от* $\frac{1}{2}$ *окажется таким же или больше?*

Решение.

1. *Зададим схему Бернулли, успех — выпадение герба,* $p = q = \frac{1}{2}$;

2.

$$A = \{ \textit{ относительное отклонение частоты выпадения герба от } \frac{1}{2}$$
$$\textit{будет не меньше, чем в эксперименте Пирсона } \}$$

Тогда событие $\overline{A}$ *будет*

$$A = \{ \text{ относительное отклонение частоты выпадения герба от } \frac{1}{2} \\ \text{будет меньше, чем в эксперименте Пирсона } \}$$

$$P\{\overline{A}\} = P\left\{\left|\frac{k}{n} - \frac{1}{2}\right| < \frac{12012}{n}\right\} = P\{11988 < k < 12012\} \approx$$
$$\approx \langle \text{интегральной теоремой Муавра-Лапласа, т. к. } n = 24000 \gg 1 \rangle \approx$$
$$\approx \Phi_0(x_2) - \Phi_0(x_1) =$$
$$= \left\langle x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{12012 - 12000}{\sqrt{24000 \cdot \frac{1}{4}}} \approx 0.155,\ x_1 = -x_2 \right\rangle =$$
$$= 2\,\Phi_0(0.155) \approx 0.123$$

Таким образом, $P(A) = 1 - P(\overline{A}) = 0.877.$

3 Математическая статистика

3.1 Основные понятия выборочной теории

3.1.1 Основные определения

Теория вероятностей является одной из ветвей «чистой» математики, которая строится дедуктивно, исходя из определённого набора аксиом. Математическая статистика является разделом прикладной математики, которая строится индуктивно от наблюдений к гипотезе. При этом выводы и аргументация математической статистики базируется на результатах теории вероятностей.

Пример 1. ***Типовые задачи теории вероятностей:*** *известно, что при одном подбрасывании симметричной монеты герб выпадает с вероятностью p. Какова вероятность того, что при n подбрасываниях герб выпадет n раз?*

Пример 2. ***Типовая задача математической статистики:*** *известно, что при n подбрасываниях монеток герб выпал n раз. Чему равна вероятность p выпадения герба при одном подбрасывании?*

Основная задача математической статистики — разработка методов получения научно обоснованных выводов о массовых явлениях или процессах по данным наблюдения или экспериментов. Эти выводы относятся не к отдельным экспериментам, а а являются утверждениями относительно вероятностных характеристик изучаемых явлений или процессов.

Замечание. *Упрощённо можно считать, что «общая» задача математической статистики формулируется следующим образом:*

Дано: *имеется случайная величина X, закон распределения которой неизвестен или известен не полностью. Требуется по результатам серии наблюдений сделать выводы о законе распределения этой случайной величины.*

При этом различают первую и вторую задачи математической статистики.

Первая основная задача Математической статистики. Есть случайная величина X, закон распределения которой неизвестен. Требуется найти закон распределения случайной величины X (требуется проверить гипотезу о законе распределения случайной величины X).

Вторая основная задача математической статистики. Есть некоторая случайная величина X, закон распределения которой известен с точностью до вектора параметров (таким образом, известен общий вид закона распределения случайной величины X, но неизвестны значения некоторых числовых параметров этого закона). Нужно найти или оценить значения этих параметров.

Пример. ***Вторая основная задача МС.*** *Известно, что случайная величина X имеет нормальное распределение, но значения параметров n и σ неизвестны. Требуется по результатам наблюдений оценить значения этих параметров. $\vec{\theta} = (m, \sigma^2)$ — вектор неизвестных параметров.*

Определение. *Множество возможных значений случайной величины X называют генеральной совокупностью.*

Определение. ***Случайной*** *выборкой из генеральной совокупности X называется случайный вектор $\vec{X} = (X_1, \ldots, X_n)$, где $X_1, \ldots, X_n$ — независимые (в совокупности) случайные величины, каждая из которых имеет то же распределение, что и X. При этом n называют объёмом случайной выборки.*

Замечание. *Всюду далее векторы будут обозначаться стрелочкой $\vec{X}$,но не чёрточкой $\overline{X}$. Черта будет обозначать выборочное среднее: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.*

Замечание 2. *Пусть $F(t)$ — функция распределения случайной величины X. Тогда функция распределения случайной выборки $\vec{X}$ объёма n из генеральной совокупности X:*

$$
\begin{gathered}
F_{\vec{X}}(t_1, \ldots, t_n) = P\{X_1 < t_1, \ldots, X_n < t_n\} = \langle X_i \text{ — \textit{независимы} } \rangle = \\
= P\{X_1 < t_1\} \cdot P\{X_2 < t_2\} \cdot \ldots \cdot P\{X_n < t_n\} = \langle X_i \sim X \rangle = \\
=? \langle \text{ \textit{далее додумано редактором} } \rangle \prod_{i=1}^{n} F(t_i)
\end{gathered}
$$

Определение. *Любую возможную реализацию $\vec{x} = (x_1, \ldots, x_n)$ случайной выборки $\vec{X}$ называются выборкой («просто» выборкой) из генеральной совокупности X. При этом число x_k называют k-ым элементом выборки $\vec{x}$.*

Определение. *Множество всех возможных значений случайной выборки $\vec{x}$ называют выборочным пространством и обозначают X_n.*

Определение. *Любую функцию $g(\vec{X})$ случайной выборки $\vec{X}$ называют статистикой или выборочной характеристикой.*

Закон распределения $F_g(\vec{t})$ называются выборочным законом распределения статистики g.

Замечание 1. *Значение $g(\vec{x})$ статистики g на выборке $\vec{x}$ называется её выборочным значением.*

Замечание 2. *Пусть $\vec{x}$ — реализация случайной выборки $\vec{X}$. При этом $\vec{x} = (x_1, \ldots, x_n)$, это позволяет считать, что случайная величина X моделируется дискретной случайной величиной, закон (ряд) распределения которой имеет следующий вид:*

$$
\begin{array}{ccccc}
\vec{x} & x_1 & x_2 & \ldots & x_n \\
P & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n}
\end{array}
$$

Математическое ожидание такой дискретной случайной величины $m = \frac{1}{n}\sum x_i$, дисперсия $\sigma^2 = \frac{1}{n}\sum(x_i - m)^2$

Эти соображения приводят к следующему определению.

Определение. *Выборочным средним (выборочным математическим ожиданием) называют статистику*

$$\hat{m}(\vec{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Определение. *Выборочной дисперсией называют вот статистику*

$$\hat{\sigma}^2(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Лекция №19, 25.02.2019

Определение. *Выборочным начальным моментом порядка k называют статистику*

$$\hat{m}_k(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Определение. *Центральным выборочным моментом порядка k называют статистику*[41]

$$\hat{\nu}_k(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

Замечание. *Очевидно, что $\overline{X} = \hat{m}_1(\vec{X})$, $\hat{\sigma}^2(\vec{X}) = \hat{\nu}_2(\vec{X})$.*

3.1.2 Предварительная обработка результатов эксперимента

Пусть $\vec{X}$ — случайная выборка из генеральной совокупности X, а $\vec{x}$ — её реализация (выборка).

3.1.2.1 Вариационный ряд

Расположим элементы выборки $\vec{x}$ в порядке неубывания, i-ый элементы полученной последовательности обозначим $x_{(i)}$, $i = \overline{1;\, n}$: $x_{(1)} \leqslant \ldots \leqslant x_{(n)}$.

Замечание. *$x_{(1)}$ — наименьший член этого ряда, $x_{(n)}$ — наибольший член этого ряда.*

Определение. *Последовательность $x_{(1)}, \ldots, x_{(n)}$ называют вариационным рядом выборки $\vec{x}$. При этом $x_{(i)}$ называют i-ым членом вариационного ряда.*

Определение. *Вариационным рядом случайной выборки $\vec{X} = (X_1, \ldots, X_n)$ называется последовательность случайных величин $X_{(1)}, \ldots, X_{(n)}$, где для каждой реализации $\vec{x}$ случайной выборки $\vec{X}$ случайная величина $X_{(i)}$ принимает значение, равное i-ому члену вариационного ряда для выборки $\vec{x}$.*

[41] ν — строчная буква «ню» греческого алфавита. — Прим. ред.

Замечание 1. $P\{X_{(i)} \leqslant X_{(i+1)}\} = 1,\ i = \overline{1;\, n-1}$

Замечание 2. *Пусть $F(x)$ — функция распределения случайной величины X; тогда*

$$F_{X_{(n)}}(x) = P\{X_{(n)} < x\} = \left\langle X_{(n)} = \max_{i=\overline{1;\,n}}\{X_i\} \right\rangle =$$
$$= P\{X_1 < x,\, X_2 < x,\, \ldots,\, X_n < x\} = \langle X_{(i)} \text{ — независимые}\rangle =$$
$$= P\{X_1 < x\} \cdot \ldots \cdot P\{X_n < x\} = F(x) \cdot \ldots \cdot F(x) = [F(x)]^n$$

$$F_{X_{(1)}} = P\{X_{(1)} < x\} = 1 - P\{X_{(1)} \geqslant x\} = \left\langle X_{(1)} = \min_{i=\overline{1;\,n}\{X_i\}} \right\rangle =$$
$$= 1 - P\{X_1 \geqslant x\} \cdot \ldots \cdot P\{X_n \geqslant x\} =$$
$$= 1 - (1 - P\{X_1 < x\}) \cdot \ldots \cdot (1 - P\{X_n < x\}) = 1 - (1 - F(x))^n$$

3.1.2.2 Статистический ряд

Среди элементов выборки $\vec{x} = (x_1,\, \ldots,\, x_n)$ некоторые могут совпадать. Это возможно в случае, когда случайная величина X является дискретной или в случае, если X — непрерывная случайная величина, но при проведении эксперимента полученные значения округлялись.

Пусть $z_{(1)}, \ldots, z_{(m)}$ — все попарно различные значения, встречающиеся в выборке $\vec{x}$, причём $z_{(1)} < \ldots < z_{(m)}$.

Определение. *Статистическим рядом, отвечающий выборке $\vec{x}$, называется таблица*

$z_{(1)}$	$\ldots$	$z_{(i)}$	$\ldots$	$z_{(m)}$
n_1	$\ldots$	n_i	$\ldots$	n_m

Здесь n_i — кол-во элементов выборки $\vec{x}$, которые имеют значение $z_{(i)}$.

Замечание 1. $n_1 + \ldots + n_m = n$;

Замечание 2. *n_i называют частотой значения $z_{(i)}$, а $\frac{n_i}{n}$ — относительной частотой этого значения.*

3.1.2.3 Эмпирическая и выборочная функции распределения

Пусть $\vec{x} = (x_1,\, \ldots,\, x_n)$ — выборка из генеральной совокупности X.

Определение. *Эмпирической функцией распределения, отвечающей выборке $\vec{x}$, называется функция $F_n(x) = \frac{h(x,\vec{x})}{n}$ где $h(x,\, \vec{x})$ — кол-во элементов выборки $\vec{x}$, которые имеют значение, меньшее x.*

Замечание 1. *$F_n(x)$ обладает всеми свойствами функции распределения:*

1. *F_n — неубывающая;*
2. $$\lim_{x\to-\infty} F_n(x) = 0,\ \lim_{x\to+\infty} F_n(x) = 1$$
3. *$F_n(x)$ — непрерывна слева в каждой точке $x \in \mathbb{R}$. При этом $F_n(x)$ кусочно-постоянна и изменяется скачками в точках $x = z_{(i)}$.*

   Если все элементы выборки $\vec{x}$ попарно различны, то

$$F_n(x) = \begin{cases} 0, & x \leqslant x_{(1)} \\ \frac{i}{n}, & x_{(i)} < x \leqslant x_{(i+1)},\ i = \overline{1;\, n-1} \\ 1, & x > x_{(n)} \end{cases}$$

Замечание 2. *Эмпирическая функция распределения позволяет интерпретировать выборку $\vec{x}$ как реализацию дискретной случайной величины $\tilde{X}$, закон распределения которой имеет следующий вид:*

$$\begin{array}{cccc} \tilde{X} & z_{(1)} & \ldots & z_{(m)} \\ P & \frac{n_1}{n} & \ldots & \frac{n_m}{n} \end{array}$$

Здесь n_i — частота значения $z_{(i)}$.

В дальнейшем это позволит рассматривать числовые характеристики случайной величины $\tilde{X}$ как приближённые значения числовых характеристик случайной величины X.

Определение. *Выборочной функцией распределения, отвечающей случайной выборке $\hat{X}$, называется функция*

$$\hat{F_n} = \frac{n(x,\, \vec{X})}{n}$$

где $n(x,\, \vec{X})$ — случайная величина, которая для каждой реализации $\vec{x}$ случайной выборки $\vec{X}$ принимает значение, равное $n(x,\, \vec{x})$.

Замечание 1. *Зафиксированное некоторое значение $x \in \mathbb{R}$,*

$$P\{X_i < x\} = \langle X_i \text{ независима и одинаково распределена с } X\rangle = \\ = \underbrace{P\{X < x\}}_{\text{обозначим } p \text{ (для данного фиксированного } x)}$$

Замечание 2. *Какие в принципе значения может принимать функция $\hat{F}_n(x)$ при фиксированном x?*

$$0,\, \frac{1}{n},\, \frac{2}{n},\, \ldots,\, \frac{n-1}{n},\, 1$$

Тогда $\hat{F}_n(x) = \frac{k}{n} \iff n(x, \vec{X}) = k$, *поэтому*

$$P\{\hat{F}_n(x) = \tfrac{k}{n}\} = P\{n(x, \vec{X}) = k\} =$$

$$= P\{\text{среди случайных величин } X_1, \ldots, X_n \text{ ровно } k \text{ приняли значение меньше, чем } x\} =$$

$$= P\{\text{в серии из } n \text{ испытаний по схеме Бернулли с вероятностью успеха } p$$
$$\text{произошло ровно } k \text{ успехов }\} = C_n^k p^k q^{n-k}$$

где $p = P\{X_i < x\}$, $q = 1 - p$.

Таким образом, для каждого фиксированного x *случайная величина* $n \cdot \hat{F}_n$ *имеет биномиальное распределение, т. е.* $n \cdot \hat{F}_n \sim B(n, p)$.

Теорема. *Для любого фиксированного* $x \in \mathbb{R}$ *последовательность* $\hat{F}_n(x)$ *сходится по вероятности к значению* $F(x)$ *теоретической (т. е. «истинной») функции распределения случайной величины* X*:*

$$\forall x \in \mathbb{R}\ \hat{F}_n(x) \xrightarrow[n\to\infty]{P} F(x)$$

Доказательство. $\hat{F}_n(x)$ — *относительная частота успеха в серии из* n *испытаний по схеме Бернулли с вероятностью успеха* p. *В соответствии с законом больших чисел в форме Бернулли* $\hat{F}_n(x) \xrightarrow[n\to\infty]{P} p$, *но* $p = P\{X < x\} = F(x)$.

3.1.2.4 Интервальный статистический ряд

Выше было введено понятие статистического ряда, однако при больших объёмах выборки ($n > 50$) его использование не очень удобно ввиду громоздкости. В этом случае отрезок $[x_{(1)}; x_{(n)}]$ разбивают на m равновеликих интервалов и для каждого интервала указывают кол-во принадлежащих ему элементов выборки. При этом, как правило, m определяют как:

$m = [\log_2 n] + 1$ — кол-во интервалов, $\triangle = \frac{x_{(n)} - x_{(1)}}{m}$ — ширина одного интервала.

Интервалы:

$$J_i = [x_{(1)} + (i-1)\triangle;\, x_{(1)} + i\triangle),\ i = \overline{1;\, m-1}$$
$$J_m = [x_{(1)} + (m-i)\triangle;\, \underbrace{x_{(1)} + m\triangle}_{x_{(n)}}]$$

Обратите внимание на границы интервалов.

Определение. *Интервальным статистическим рядом называют таблицу*

$$\begin{matrix} J_1 & \ldots & J_i & \ldots & J_m \\ n_1 & \ldots & n_i & \ldots & n_m \end{matrix}$$

Здесь n_i — *количество элементов выборки* $\vec{x}$, *принадлежащих* J_i.

3.1.2.5 Эмпирическая плотность и гистограмма

Определение. *Эмпирической плотностью, отвечающей выборке $\vec{x}$, называют функцию*

$$f_n(x) = \begin{cases} \frac{n_i}{n\triangle}, & x \in J_i \\ 0 & \textit{иначе} \end{cases}$$

Здесь $\triangle$, J_i выбираются из интервального ряда.

Замечание.

$$\int\limits_{-\infty}^{+\infty} f_n(x)\,dx = \sum_{i=1}^{m} (\textit{площадей прямоугольников}) = \sum_{i=1}^{m} \frac{n_i}{n \cdot \cancel{\triangle}} \cdot \cancel{\triangle} = \frac{1}{n} \cancelto{n}{\sum_{i=1}^{m} n_i} = 1$$

Эмпирическая функция плотности обладает всеми свойствами «обычной» функции плотности и по этой причине является её статистическим аналогом.

Аналогично случаю выборочной функции распределения можно показать, что при $n \to \infty$

$$f_n(x) \to f(x)$$

где f — функция плотности случайной величины X (если X — непрерывная).

Определение. *График эмпирической функции плотности называется гистограммой.*

3.1.2.6 Полигон частот

Предположим, что для данной выборки $\vec{x}$ построена гистограмма.

Определение. *Полигоном частот для выборки $\vec{x}$ называется ломаная, звенья которой соединяют середины верхних сторон прямоугольников гистограммы.*

Лекция №20, 04.03.2019

3.2 Основные распределения, используемые в математической статистике

3.2.1 Гамма-функция Эйлера

Определение. *Гамма-функцией Эйлера называется функция вида*

$$\Gamma(\alpha) = \int\limits_{0}^{+\infty} x^{\alpha-1}\, e^{-x}\, dx \tag{14}$$

Будем рассматривать Γ-функцию для вещественных значений α.

Интеграл в правой части 14 является несобственным и сходится при $\alpha > 0$.

3.2.1.1 Некоторые свойства гамма-функции

1. $\Gamma(\alpha)$ является бесконечно число раз дифференцируемой функцией при $\alpha \in (0;\,+\infty)$, причём

$$\Gamma^k(\alpha) = \int\limits_0^{+\infty} (\ln x)^k x^{\alpha-1} e^{-x}\,dx$$

2.

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha),\ \alpha > 0$$

3.

$$\Gamma(1) = \int\limits_0^{+\infty} x^{1-1}e^{-x}dx = \int\limits_0^{+\infty} e^{-x}dx = 1$$

4. Из свойств 2, 3:

$$\Gamma(n+1) = n!,\ n \in \mathbb{N}_0$$

Замечание. *В силу свойств 2 и 4 часто говорят, что* Γ*-функция является обобщением понятия факториала на случай вещественных значений аргумента.*

5.

$$\Gamma(\tfrac{1}{2}) = \int\limits_0^{+\infty} x^{\frac{1}{2}-1}e^{-x}\,dx = \int\limits_0^{+\infty} \frac{1}{\sqrt{x}}e^{-x}\,dx = 2\int\limits_0^{+\infty} e^{-x}d\sqrt{x} = \langle \sqrt{x} = t\rangle =$$
$$= 2\int\limits_0^{+\infty} e^{-t^2}\,dt = \sqrt{\pi}$$

6.

$$\Gamma(n+\tfrac{1}{2}) = \langle\text{свойство 2}\rangle = (n-\tfrac{1}{2})\cdot\Gamma(n-\tfrac{1}{2}) = \langle\text{свойство 2}\rangle =$$
$$= (n-\tfrac{1}{2})\cdot(n-\tfrac{3}{2})\cdot\Gamma(n-\tfrac{3}{2}) =$$
$$= (n-\tfrac{1}{2})\cdot(n-\tfrac{3}{2})\cdot\ldots\cdot\tfrac{1}{2}\cdot\cancelto{\sqrt{\pi}}{\Gamma(\tfrac{1}{2})} = \tfrac{(n-1)\cdot(n-3)\cdot\ldots\cdot 3\cdot 1}{2^n}\sqrt{\pi},\ n \in \mathbb{N}_0$$

7. Примерный график функции $\Gamma(x)$...

...находится вот здесь. Смотрите только значения $x > 0$.

3.2.2 Гамма-распределение

Определение. *Говорят, что случайная величина* ξ *имеет гамма-распределение с параметрами* λ *и* α*, если её функция плотности имеет вид*

$$f_\xi(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

Обозначается как $\xi \sim \Gamma(\lambda, \alpha)$.

Замечание. *Экспоненциальное распределение*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

является частным случаем гамма-распределения для значения $\alpha = 1$*, т. е.* $Exp(\lambda) = \Gamma(\lambda, 1)$.

Теорема. *Пусть*

1. $\xi_1 \sim \Gamma(\lambda, \alpha_1)$;
2. $\xi_2 \sim \Gamma(\lambda, \alpha_2)$;
3. ξ_1, ξ_2 *— независимые.*

   Тогда

$$\xi_1 + \xi_2 \sim \Gamma(\lambda, \alpha_1 + \alpha_2)$$

Следствие. *Если случайные величины* $\xi_i \sim \Gamma(\lambda, \alpha_i)$, $i = \overline{1; n}$*, то*

$$\xi_1 + \ldots + \xi_n \sim \Gamma(\lambda, \alpha_1 + \ldots + \alpha_2)$$

3.2.3 Распределение Рэлея

Пусть

1. $\xi \sim N(0, \sigma^2)$ — нормальная случайная величина с нулевым математическим ожиданием;
2. $\eta = \xi^2$. ([42])

Определение. *В этом случае говорят, что* η *имеет распределение Рэлея с параметром* σ*. Можно показать, что в этом случае*

$$f_\eta(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}\sigma} e^{-\frac{x}{2\sigma^2}}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

Замечание. *Распределение Рэлея является частным случаем гамма-распределения для* $\lambda = \frac{1}{2\sigma^2}$, $\alpha = \frac{1}{2}$*, т. е.* $\eta \sim \Gamma(\frac{1}{2\sigma^2}, \frac{1}{2})$.

[42] η — строчная буква «э́та» греческого алфавита. — Прим. ред.

3.2.4 Распределение хи-квадрат

Пусть

1. $\xi_1, \ldots, \xi_n$ — независимые случайные величины;
2. $\xi_i \sim N(0,\, 1), i = \overline{1;\, n}$;
3. $\eta = \xi_1^2 + \ldots + \xi_n^2$.

Определение. *В этом случае говорят, что случайная величина η имеет распределение хи-квадрат с n степенями свободы.*

Обозначение $\eta \sim \chi^2(n)$. ([43])

Замечание 1. *Если $\xi_i \sim N(0,\, 1)$, то $\xi_i^2 \sim \Gamma(\frac{1}{2},\, \frac{1}{2})$. Тогда по свойству Γ-распределения $\eta = \xi_1^2 + \ldots + \xi_n^2 \sim \Gamma(\frac{1}{2},\, \frac{n}{2})$.*

Замечание 2. *Если случайные величины $\eta_1, \ldots, \eta_m$ независимы и $\eta_i \sim \chi^2(n_i),\ i = \overline{1;\, m}$, то $\eta_1 + \ldots + \eta_m \sim \Gamma(\frac{1}{2};\, \frac{n_1 + \ldots + n_m}{2}) = \chi^2(n_1 + \ldots + n_m)$.*

Замечание 3. *Пусть $\eta \sim \chi^2(n)$. Тогда график функции плотности случайной величины η выглядит примерно так (см. случай для $n \gg 4$).*

3.2.5 Распределение Фи́шера

Пусть

1. ξ_1 и ξ_2 — независимы;
2. $\xi_i \sim \chi^2(n_i),\ i = \overline{1;\, 2}$;
3. $\zeta = \frac{n_2 \xi_1}{n_1 \xi_2}$. ([44])

Определение. *В этом случае говорят, что случайная величина ζ имеет распределение Фи́шера со степенями свободы n_1 и n_2.*

Обозначение $\zeta \sim F(n_1,\, n_2)$.

Замечание 1. *Если $\zeta \sim F(n_1,\, n_2)$, то $\frac{1}{\zeta} \sim F(n_2,\, n_1)$.*

Замечание 2. *Можно показать, что*

$$f_\zeta(x) = \begin{cases} c \cdot \dfrac{x^{\frac{n_1}{2} - 1}}{(1 + \frac{n_1}{n_2} x)^{\frac{n_1 + n_2}{2}}}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

[43] χ — строчная буква «хи» греческого алфавита. — Прим. ред.

[44] ζ — строчная буква «дзета» греческого алфавита. — Прим. ред.

где

$$c = \frac{\left(\frac{n_1}{n_2}\right)^{\left(\frac{n_1}{2}\right)}}{B\left(\frac{n_1}{2},\, \frac{n_2}{2}\right)}$$

где

$$B(p,\, q) = \int\limits_0^1 x^{p-1}(1-x)^{q-1}\, dx$$

— бета-функция Эйлера.

3.2.6 Распределение Стью́дента

Пусть

1. $\xi_1,\, \xi_2$ — независимые случайные величины;
2. $\xi_1 \sim N(0,\, 1),\ \xi_2 \sim \chi^2(n)$;
3. $\zeta = \frac{\xi_1}{\sqrt{\xi_2}}\sqrt{n}$.

Определение. *В этом случае говорят, что случайная величина ζ имеет распределение Стью́дента с n степенями свободы.*

Обозначение $\zeta \sim St(n-1)$.

Замечание 1. *Можно показать, что*

$$f_\zeta(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x \in \mathbb{R}$$

Замечание 2. *Если $\zeta_1 \sim St(n_1)$, $\zeta_2 \sim St(n_2)$, где $n_1 < n_2$, то графики имеют вид...*
...этот.

3.3 Точечные оценки

3.3.1 Основные понятия

Задача. ***Вторая задача математической статистики.***
Пусть X — случайная величина, общий вид закона распределения которой известен, но неизвестны значения одного или нескольких параметров этого закона. Требуется найти (оценить) значения этих параметров.

Для простоты будем считать, что функция распределения случайной величины X зависит от одного неизвестного параметра $\theta \in \mathbb{R}$, т. е. $F(x,\, \theta)$ — функция распределения случайной величины X.

Определение. *Статистику $\hat{\theta}(\vec{X})$ называют точечной оценкой параметра θ, если её выборочное значение принимается в качестве значения параметра θ, т. е.*

$$\theta := \hat{\theta}(\vec{x})$$

Пример. *Пусть X — случайная величина, $m = MX$ и оно неизвестно; рассмотрим следующие точечные оценки для m:*

$$\hat{m}_1(\vec{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{m}_2(\vec{X}) = \tfrac{1}{2}[X_{(1)} + X_{(n)}]$$

$$\hat{m}_3(\vec{X}) = \begin{cases} X(\frac{n+1}{2}), & \text{если } n \text{ — нечётное,} \\ \frac{1}{2}[X(\frac{n}{2}) + X(\frac{n}{2}+1)] & \text{иначе} \end{cases}$$

($\hat{m}_3(\vec{X})$ — средний член вариационного ряда)

$$\hat{m}_4(\vec{X}) = X_1$$

($\hat{m}_4(\vec{X})$ — результат первого наблюдения)

$$\hat{m}_5(\vec{X}) = e^{(\overline{X})^2} \cdot \sin(X_1^3 + 2 \cdot X_2^3 + \ldots + n \cdot X_n^3)$$

Понятно, что некоторые из этих оценок более удачны, некоторые менее, некоторые абсурдны. Для сравнения качества точечных оценок используются следующие их свойства:

1. *Несмещённость;*
2. *Состоятельность;*
3. *Эффективность.*

Лекция №21, 11.03.2019

Замечание. *Будет ли некоторая статистика $\hat{\theta}(\vec{X})$ служить точечной оценкой для параметра θ, зависит от нашего желания: если мы используем $\hat{\theta}(\vec{x})$ в качестве значения параметра θ, то эта статистика будет являться точечной оценкой для θ. В противном случае — не будет.*

3.3.2 Несмещённость точечной оценки

Пусть

1. $F(x, \theta)$ — функция распределения случайной величины X с неизвестным параметром θ;
2. $\hat{\theta}(\vec{X})$ — точечная оценка для θ.

Определение. *Точечная оценка $\hat{\theta}(\vec{X})$ называется несмещённой оценкой для θ, если $\exists M[\hat{\theta}(\vec{X})] = \theta$.*

Пример 1. *Рассмотрим оценки математического ожидания точечных оценок из предыдущего примера:*

$$M[\hat{m}_1(\vec{X})] = M[\tfrac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} MX_i = \frac{1}{n}\sum_{i=1}^{n} m = m$$

т. е. $\hat{m}_1$ — несмещённая оценка для m.

$$M[\hat{m}_4(\vec{X})] = M[X_1] = \langle X_1 \sim X\rangle = m$$

т. е. $\hat{m}_4$ — несмещённая оценка для m.

Пример 2. *Пусть X — случайная величина, также $\exists DX = \sigma^2$ и она неизвестна. Можно показать, что выборочная дисперсия $\hat{\sigma}^2(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ является смещённой оценкой дисперсии: $M[\hat{\sigma}^2(\vec{X})] = \frac{n-1}{n}\sigma^2$ (доказано на семинаре).*

По этой причине рассмотрим статистику

$$S^2(\vec{X}) = \frac{n}{n-1}\hat{\sigma}^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

для которой

$$M[S^2(\vec{X})] = \frac{n}{n-1}M[\hat{\sigma}^2] = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

Таким образом, $S^2(\vec{X})$ является несмещённой оценкой дисперсии.

Определение. *Статистика $S^2(\vec{X})$ называется исправленной выборочной дисперсией.*

3.3.3 Состоятельность точечной оценки

Пусть X — случайная величина, $F(x, \theta)$ — функция распределения случайной величины X, θ — неизвестный параметр, $\hat{\theta}(\vec{X})$ — точечная оценка для θ.

Определение. *Оценка $\hat{\theta}(\vec{X})$ называется состоятельной оценкой параметра θ, если*

$$\hat{\theta}(\vec{X}) \xrightarrow[n\to\infty]{P} \theta$$

где n — объём выборки $\vec{X}$.

Замечание 1. *Условие из определения точечной оценки можно записать в виде: $\forall\epsilon > 0 \lim\limits_{n\to\infty} P\{|\hat{\theta}(\vec{X}) - \theta| - \epsilon\} = 1$*

Замечание 2. *Если оценка $\hat{\theta}(\vec{X})$ не является состоятельной, то она никому не нужна.*

Пример. *Пусть*

1. *X — случайная величина;*
2. *$\exists MX = m$;*
3. *$\exists DX = \sigma^2$.*

Покажем, что выборочное среднее $\overline{X}$ является состоятельной оценкой для m.

1. *Последовательность X_1, X_2, …, X_n, … является последовательностью независимых одинаково распределённых случайных величин;*
2. *$\exists MX_i = m$, $\exists DX_i = \sigma^2$, $i \in \mathbb{N}$;*
3. *Из 1, 2 следует, что последовательность X_1, X_2 и т. д. удовлетворяет закону больших чисел в форме Чебышева, из чего следует*

$$\forall\epsilon > 0 P\{|\overline{X} - m| < \epsilon\} \xrightarrow[n\to\infty]{} 1$$

т. е. $\overline{X} \xrightarrow[n\to\infty]{P} m$.

Замечание 1. *Состоятельность оценки $\overline{X}$ для m можно доказать и не предполагая существование конечной дисперсии.*

Замечание 2. *Можно показать, что статистика $S^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ является состоятельной оценкой дисперсии случайной величины X, если эта случайная величина имеет момент до четвёртого порядка включительно.*

Замечание 3. *Более общо: если случайная величина X имеет начальные и центральные моменты, то их выборочные аналоги являются их состоятельными оценками. При этом все выборочные моменты порядка два и выше являются смещёнными оценками соответствующих теоретических моментов.*

Пример. *Пусть*

1. $X \sim N(m, \sigma^2)$, *причём* m, σ^2 *— неизвестны;*
2. $\hat{m}(\vec{X}) = X_1$ *— результат первого наблюдения — точечная оценка для* m.

Покажем, что $\hat{m}$ *не является состоятельной оценкой. Зафиксируем* $\epsilon > 0$:

$$P\{|\hat{m}(\vec{X}) - m| < \epsilon\} = P\{|X_1 - m| < \epsilon\} =$$
$$= P\{m - \epsilon < X_1 < m + \epsilon\} = \langle X_1 \sim X \sim N(m, \sigma^2)\rangle =$$
$$= \Phi_0(\tfrac{m+\epsilon-m}{\sigma}) - \Phi_0(\tfrac{m-\epsilon-m}{\sigma}) = 2\Phi_0(\tfrac{\epsilon}{\sigma}) \neq 1, \text{ если } \frac{\epsilon}{\sigma} \neq +\infty$$

Поэтому $P\{|\hat{m}(\vec{X}) - m| < \epsilon\} \xrightarrow[n\to\infty]{} 1$.

3.3.4 Эффективность точечной оценки

1. Дисперсия характеризует разброс значений случайной величины относительно её математического ожидания: чем больше дисперсия, тем больше разброс.

   Если Y, Z — две случайные величины, $MY = MZ = m$, но если $DY < DZ$, то качественно эту ситуацию можно проиллюстрировать следующим образом:

   <Здесь был рисунок, но в нём не было никакого откровения.>

2. Пусть $\hat{\theta}(\vec{X})$ — несмещённая оценка параметра θ. $\vec{X}$ — случайный вектор, $\hat{\theta}(\vec{X})$ — скалярная функция случайного вектора, т. е. $\hat{\theta}(\vec{X})$ является случайной величиной. Возможно, что значения $\hat{\theta}(\vec{X})$ кучно группируются около математического ожидания. Возможно, что менее кучно. С точки зрения качества точечной оценки первая ситуация предпочтительнее.

Пусть $\hat{\theta}_1(\vec{X})$, $\hat{\theta}_2(\vec{X})$ — несмещённые оценки для параметра θ.

Определение. *Оценка* $\hat{\theta}_1$ *называется более эффективной, чем оценка* $\hat{\theta}_2$, *если*

1. $\exists D\hat{\theta}_1$, $\exists D\hat{\theta}_2$.
2. $D\hat{\theta}_1 < D\hat{\theta}_2$.

Определение. *Оценка* $\hat{\theta}$ *называется эффективной оценкой для параметра* θ, *если*

1. $\hat{\theta}$ *является несмещённой оценкой для* θ;

2. *$\hat{\theta}$ обладает наименьшей дисперсией среди всех несмещённых оценок для θ.*

Замечание. *Иногда говорят об эффективной оценке в классе оценок Θ. Если Θ — некоторое множество несмещённых оценок для параметра θ, то оценка $\hat{\theta} \in \Theta$ называется эффективной оценкой для θ в классе Θ, если $\hat{\theta}$ обладает наименьшей дисперсией среди всех оценок класса Θ, т. е. $\forall \tilde{\theta} \in \Theta \ D\hat{\theta} \leqslant \tilde{\theta}$.*

Пример. *Пусть*

1. *X — случайная величина;*

2. *$\exists MX = m$, $\exists DX = \sigma^2$.*

Покажем, что выборочное среднее $\overline{X}$ является эффективной оценкой для m в классе линейных оценок.

1. *Линейная оценка имеет вид:*

$$\hat{m}(\vec{X}) = \lambda_1 X_1 + \ldots + \lambda_n X_n$$

где $\lambda_j \in \mathbb{R}$, $j = \overline{1;\, n}$.

2. *Т. к. оценка должна быть несмещённой, то*

$$M[\hat{m}(\vec{X})] = M\left[\sum_{i=1}^{n} \lambda_i X_i\right] = \sum_{i=1}^{n} \lambda_i M X_i = \langle X_i \sim X,\ i = \overline{1;\, n} \rangle =$$
$$= \sum_{i=1}^{n} \lambda_i m = m \sum_{i=1}^{n} \lambda_i$$

Требуем: $M[\hat{m}(\vec{X})] = m \implies \sum\limits_{i=1}^{n} \lambda_i = 1$.

3. *Подберём в линейной оценке $\hat{m}(\vec{X}) = \sum\limits_{i=1}^{n} \lambda_i X_i$ коэффициенты λ_i таким образом, чтобы $D[\hat{m}(\vec{X})]$ была линейной среди значений дисперсии всевозможных линейных оценок.*

$$D[\hat{m}] = D[\sum_{i=1}^{n} \lambda_i X_i] = \langle X_i \text{ — независимы} \rangle = \sum_{i=1}^{n} \lambda_i^2 D X_i = \sigma^2 \sum_{i=1}^{n} \lambda_i^2$$

Таким образом, приходим к значению поиска условия экстремума:

$$\begin{cases} f(\lambda_1,\, \ldots,\, \lambda_n) = \sum\limits_{i=1}^{n} \lambda_i^2 \longrightarrow \min \\ \sum\limits_{i=1}^{n} \lambda_i = 1 \end{cases}$$

Замечание. *Если бы экстремум был безусловный, то необходимые условия экстремума:*

$$\begin{cases} f(x_1, \ldots, x_n) \to min \\ \frac{\delta f}{\delta x_i} = 0, \ i = \overline{1; n} \end{cases}$$

Если условный:

$$\begin{cases} f(x_1, \ldots, x_n) \to min \\ \varphi(x_1, \ldots, x_n) = 0 \end{cases}$$

Составим функцию Лагранжа:

$$L(\lambda_1, \ldots, \lambda_n, \mu) = f(\lambda_1, \ldots, \lambda_n) - \mu(\sum_{i=1}^{n} \lambda_i - 1) = (\lambda_1^2 + \ldots + \lambda_n^2) - \mu(\lambda_1 + \ldots + \lambda_{n-1})$$

Необходимые условия экстремума:

$$\begin{cases} \frac{\delta L}{\delta \lambda_1} = 2\lambda_1 - \mu = 0 \\ \ldots \\ \frac{\delta L}{\delta \lambda_n} = 2\lambda_n - \mu = 0 \\ \frac{\delta L}{\delta \mu} = -(\sum\limits_{i=1}^{n} \lambda_i - 1) = 0 \end{cases}$$

Из этих условий (кроме последнего) следует $\lambda_i = \frac{\mu}{2}, \ i = \overline{1; n}$.

Используем последнее условие:

$$\sum_{i=1}^{n} \underbrace{\frac{\mu}{2}}_{} \lambda_i = 1$$

$\frac{\mu n}{2} = 1, \ \mu = \frac{2}{n}$.

Тогда $\lambda_i = \frac{1}{n}, \ i = \overline{1; n}$.

Можно, проверив достаточное условие экстремума, показать, что точка $(\frac{1}{n}, \ldots, \frac{1}{n}$ *является условием минимума функции* $f(\lambda_1, \ldots, \lambda_n)$. *Таким образом, линейная оценка с наименьшей дисперсией:*

$$\hat{m}(\vec{X}) = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

Соответствующее значение дисперсии:

$$D[\hat{m}(\vec{X})]\big|_{(\lambda_1, \ldots, \lambda_n) = (\frac{1}{n}, \ldots, \frac{1}{n})} = \sigma^2 (\sum_{i=1}^{n} \lambda_i)\big|_{\lambda_i \equiv \frac{1}{n}} = \sigma^2 \sum_{i=1}^{n} \frac{1}{n^2} = \frac{\sigma^2}{n}$$

Теорема. ***О единственности эффективной оценки.***

Пусть $\hat{\theta}_1(\vec{X})$ и $\hat{\theta}_2(\vec{X})$ — две эффективных оценки для параметра θ. Тогда

$$\hat{\theta}_1(\vec{X}) = \hat{\theta}_2(\vec{X})$$

Это равенство является равенством для двух случайных величин $\hat{\theta}_1$ и $\hat{\theta}_2$, поэтому оно понимается в вероятностном смысле:

$$P\{\vec{X} \in \{\vec{x}\colon \hat{\theta}_1(\vec{x}) \neq \hat{\theta}_2(\vec{x})\}\} = 0$$

Доказательство.

1. *Рассмотрим оценку $\hat{\theta} = \frac{1}{2}[\hat{\theta}_1 + \hat{\theta}_2]$.*

$$M\hat{\theta} = M[\tfrac{1}{2}(\hat{\theta}_1 + \hat{\theta}_2)] = \tfrac{1}{2}[M\hat{\theta}_1 + M\hat{\theta}_2] =$$
$$= \langle\, \hat{\theta}_1 \text{ и } \hat{\theta}_2 \text{ — эффективные и, следовательно, несмещённые } \rangle = \tfrac{1}{2}[\theta + \theta] = \theta$$

т. е. $\hat{\theta}$ также является несмещённой оценкой для θ.

$$D\hat{\theta} = \tfrac{1}{4}D[\hat{\theta}_1 + \hat{\theta}_2] = \tfrac{1}{4}[D\hat{\theta}_1 + D\hat{\theta}_2 + 2\operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2)] =$$

$\langle$ Обозначим: $D\hat{\theta}_1 = a^2 = D\hat{\theta}_2$ (из эффективности оценки). Т. к. $\hat{\theta}_1$ и $\hat{\theta}_2$ — эффективные, то они имеют одинаковую дисперсию, которая является наименьшей среди дисперсий всех несмещённых оценок для θ $\rangle$

$$= \tfrac{1}{2}[a^2 + \operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2)] \tag{15}$$

2.

$$|\operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2)| \leqslant \sqrt{\underbrace{D\hat{\theta}_1}_{=a^2} \cdot \underbrace{D\hat{\theta}_2}_{=a^2}} = a^2$$

Таким образом,

$$D\hat{\theta} \leqslant \langle \text{ см. 15 } \rangle \leqslant \tfrac{1}{2}[a^2 + a^2] = a^2 \tag{16}$$

$\hat{\theta}$ — несмещённая оценка для θ, а $\hat{\theta}_1$, $\hat{\theta}_2$ — эффективные оценки $\implies$ $D\hat{\theta}_1 = D\hat{\theta}_2 = a^2 \leqslant D\hat{\theta}$.

С учётом 16: $D\theta = a^2$.

3. *15 $\implies$*

$$\underbrace{a^2}_{D\hat{\theta}} = \tfrac{1}{2}[a^2 + \operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2)] \implies \operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2) = a^2$$

т. е.

$$\operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2) = \sqrt{D\hat{\theta}_1 \cdot D\hat{\theta}_2}$$

Следовательно, по свойствам ковариации, $\hat{\theta}_1$ и $\hat{\theta}_2$ связаны положительной линейной зависимостью, т. е. ($k > 0$)

$$\hat{\theta}_1 = k\hat{\theta}_2 + b \tag{17}$$

4. 17 $\Longrightarrow$

$$\underbrace{D\hat{\theta}_1}_{=a^2} = k^2 \underbrace{D\hat{\theta}_2}_{=a^2} \Longrightarrow k^2 = 1 \Longrightarrow k = 1$$

Тогда $\hat{\theta}_1 = \hat{\theta}_2 + b \Longrightarrow \underbrace{M\hat{\theta}_1}_{=\theta} = \underbrace{M\hat{\theta}_2}_{=\theta} + b \Longrightarrow b = 0$.

Таким образом, 17 $\Longrightarrow \hat{\theta}_1 = \hat{\theta}_2$.

Лекция №22, 18.03.2019

Вопрос. *Из предыдущих рассуждений следует, чем меньше дисперсия точечной оценки, тем более эффективной будет её практическое использование. Можно ли сделать дисперсию несмещённой точечной оценки неизвестного параметра некоторого закона распределения сколь угодно малой?*

Оказывается, значения дисперсии всех несмещённых точечных оценок данного параметра (при фиксированном объёме выборки) ограничены снизу (см. неравнество Páо-Крамéра).

Пусть

1. X — непрерывная случайная величина;
2. $f(t, \theta)$ — функция плотности распределения вероятностей случайной величины X;
3. $\vec{X} = (X_1,\, \ldots,\, X_n)$ — случайная выборка из генеральной совокупности X.

Тогда функция плотности распределения случайного вектора $\vec{X}$:

$$f_{\vec{X}}(t_1,\, \ldots,\, t_n,\, \theta) = f(t_1,\, \theta) \cdot \ldots f(t_n,\, \theta)$$

(поскольку компоненты случайного вектора независимы)

Обозначим: $(t_1,\, \ldots,\, t_n) = \vec{T}$.

Определение. *Величина*

$$I(\theta) = M\left\{\left[\frac{\delta \ln f_{\vec{X}}(\vec{T},\theta)}{\delta\theta}\right]^2\right\}$$

называется количеством информации по Фишеру (в серии из n наблюдений).

Замечание. *Ниже иногда будет нужно дифференцировать по параметру под знаком интеграла:*

$$\frac{\delta}{\delta\theta}\int \varphi(\vec{T},\,\theta)\,d\vec{T} = \int \frac{\delta\varphi(\vec{T},\,\theta)}{\delta\theta}\,d\vec{T}$$

Параметрические модели, для которых справедлив данный переход будем называть регулярными.

Теорема. *Неравнество Ра́о-Краме́ра.* *Пусть*

1. *Рассматривается регулярная модель;*
2. $\hat{\theta}(\vec{X})$ *— несмещённая точечная оценка параметра θ закона распределения случайной величины X.*

Тогда

$$D\hat{\theta}(\vec{X}) \geqslant \frac{1}{I(\theta)}$$

где $I(\theta)$ — количество информации по Фишеру.

Доказательство.

1. *Обозначим $G = \{t \in \mathbb{R}\colon f(t,\,\theta) > 0$. Тогда*

$$\int\limits_{\mathbb{R}^n} f_{\vec{X}}(\vec{T},\,\theta)\,d\vec{T} = \int\limits_{G^n} f_{\vec{X}}(\vec{T},\,\theta)\,d\vec{T} = 1 \text{ (по условию нормировки)}$$

2. *Продифференцируем подчёркнутое равенство по θ: Правая часть: $\frac{\delta 1}{\delta\theta} = 0$; левая часть*

$$\frac{\delta}{\delta\theta}\int\limits_{G^n} f_{\vec{X}}(\vec{T},\,\theta)\,d\vec{T} = \langle\text{модель явл. рес.?}\rangle =$$

$$= \int\limits_{G^n} \frac{\delta f_{\vec{T}}(\vec{T},\,\theta)}{\delta\theta}\,dT = \langle\frac{\delta \ln y}{\delta\theta} = \frac{1}{y}\cdot\frac{\delta y}{\delta\theta} \implies \frac{\delta y}{\delta\theta} = y\frac{\delta \ln y}{\delta\theta}\rangle =$$

$$= \int\limits_{G^n} \frac{\delta \ln \overbrace{f_{\vec{X}}(\vec{T},\,\theta)}^{y}}{\delta\theta}\underbrace{f_X(\vec{T},\,\theta)}_{y}\,d\vec{T} = \langle M[\varphi(\vec{X})] = \int\limits_{R^n}\varphi(\vec{X})f(\vec{X})d\vec{X}\rangle =$$

$$= M\left[\frac{\delta \ln f_X(\vec{X},\,\theta)}{\delta\theta}\right]$$

Таким образом,

$$M\left[\frac{\delta \ln f_X(\vec{X},\,\theta)}{\delta\theta}\right] = 0 \tag{18}$$

3. *Т. к. $\hat{\theta}(\vec{X})$ — несмещённая оценка для θ, то*

$$\theta = M[\hat{\theta}(\vec{X})] = \langle M[\varphi(\vec{X})] = \int\limits_{\mathbb{R}^n} \varpi(\vec{X}) f_{\vec{X}}(\vec{x})\,d\vec{x}\rangle = \int\limits_{G^n} \hat{\theta}(\vec{T}) \cdot f_{\vec{X}}(\vec{T},\,\vec{\theta})\,d\vec{T}$$

Продифференцируем полученное равенство по θ: для левой части $\frac{\delta\theta}{\delta\theta} = 1$; правая часть:

$$\frac{\delta}{\delta\theta}\int\limits_{G^n} \hat{\theta}(\vec{T})\,f(\vec{T},\,\theta)\,d\vec{T} = \langle \text{ модель регулярная } \rangle = \int\limits_{G^n} \hat{\theta}(\vec{T})\frac{\delta f_{\vec{X}}(\vec{T},\,\theta)}{\delta\theta}\,d\vec{T} =$$

$$= \left\langle \frac{\delta y}{\delta\theta} = y\frac{\delta \ln y}{\delta\theta} \right\rangle = \int\limits_{G^n} \hat{\theta}(\vec{T})\frac{\delta \ln f_{\vec{X}}(\vec{T},\,\theta)}{\delta\theta} \cdot f_{\vec{X}}(\vec{T},\,\theta)\,d\vec{T} = M\left[\hat{\theta}(\vec{X}) \cdot \frac{\delta \ln f_{\vec{X}}(\vec{X},\,\theta)}{\delta\theta}\right]$$

Таким образом,

$$M[\hat{\theta}(\vec{X}) \cdot \frac{\delta \ln f_{\vec{X}}(\vec{X},\,\theta)}{\delta\theta}] = 1 \tag{19}$$

4. *Умножим обе части 18 на θ:*

$$M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\,\theta)}{\delta\theta} \cdot \theta] = 0 \tag{20}$$

Вычтем из 19 равенство 20:

$$M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\,\theta)}{\delta\theta}(\hat{\theta}(\vec{X}) - \theta)] = 1$$

Возведём обе части этого равенства в квадрат:

$$1 = \left\{ M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}(\hat{\theta}(\vec{X}) - \theta)] \right\}^2 =$$

$$= \left\{ \int\limits_{G^n} \underbrace{\frac{\delta \ln f_{\vec{X}}(\vec{T},\, \theta)}{\delta\theta}}_{a(\vec{T})} \cdot \underbrace{(\hat{\theta}(\vec{T}) - \theta)}_{b(\vec{T})} \cdot \underbrace{f_X(\vec{T},\, \theta)}_{\textit{(вес)}}\, d\vec{T} \right\}^2 =$$

$$= \Big\langle \textit{Неравенство Коши-Буняковского: } |(a,\, b)| \leqslant |a| \cdot |b|,\ \textit{где } |x| = \sqrt{(x,\, x)};$$

В пространстве $L^2(G^n)$ скалярное произведение может быть задано в виде:

$$(a,\, b) = \int\limits_{G^n} a(\vec{T}) b(\vec{T}) p(\vec{T}) d\vec{T}$$

$$\textit{где } p(\vec{T}) > 0 \textit{ в } G^n, \textit{ где } p(\vec{T}) - \textit{весовая функция} \Big\rangle =$$

$$= \{(a(\vec{T}),\, b(\vec{T})\}^2 \leqslant \langle \textit{неравенство Коши-Буняковского} \rangle \leqslant$$

$$\leqslant \underbrace{(a(\vec{T}),\, a(\vec{T}))}_{|a|^2} \cdot \underbrace{(b(\vec{T}),\, b(\vec{T}))}_{|b|^2} =$$

$$= \underbrace{\int\limits_{G^n} [\frac{\delta \ln f_{\vec{X}}(\vec{T},\, \theta)}{\delta\theta}]^2 f_{\vec{X}}(\vec{T},\, \theta) d\vec{T}}_{|a|^2} \cdot \underbrace{\int\limits_{G^n} (\hat{\theta}(\vec{T}) - \theta)^2 f_{\vec{X}}(\vec{T},\, \theta) d\vec{T}}_{|b|^2} =$$

$$= M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}]^2] \cdot M[(\hat{\theta}(\vec{T}) - \overset{M[\hat{\theta}]}{\theta})^2] = I(\theta) \cdot D[\hat{\theta}]$$

Таким образом,

$$1 \leqslant I(\theta) D(\hat{\theta}) \implies D(\hat{\theta}) \geqslant \frac{1}{I(\theta)}$$

Замечание 1. *Неравенство Рао-Крамера даёт нижнюю границу для значений дисперсии всевозможных точечных оценок параметра θ.*

Замечание 2. *Величина*

$$e(\hat{\theta}) = \frac{1}{I(\theta) \cdot D(\hat{\theta})}$$

называется показателем эффективности по Рао-Крамеру оценки $\hat{\theta}$. Из неравенства Рао-Крамера следует, что

$$0 < e(\hat{\theta}) \leqslant 1$$

При этом

$$e(\hat{\theta}) = 1 \iff D\hat{\theta} = \frac{1}{I(\theta)}$$

т. е. $D\hat{\theta}$ имеет наименьшее возможное значение. В этом случае оценка $\hat{\theta}$ называется эффективной по Рао-Крамеру.

Вообще говоря, никто не обещал, что для неизвестного параметра θ каждого из возможных законов распределения генеральной совокупности найдётся несмещённая оценка $\hat{\theta}$, для которой неравенство Рао-Крамера будет выполняться в форме равенства (т. е. никто не обещал, что нижняя графница значения дисперсии всегда достижима). Если для некоторой оценки $\hat{\theta}$ верно $D(\hat{\theta}) = \frac{1}{I(\theta)}$, то эта оценка будет эффективной по Рао-Крамеру и, следовательно, «просто» эффективной.

Замечание 3. *В каких случаях неравенство Рао-Крамера выполняется в форме равенства? Из курса линейной алгебры известно, что неравенство К.-Б.? выполняется в форме равенства TTT?, когда сомножители a, b линейно зависимы, т. е.*

$$e(\hat{\theta}) \iff \langle \textit{см. доказательство Рао-Крамера} \rangle \iff a(\vec{T}) =$$

$$= k \cdot b(\vec{T}) \iff \underbrace{\frac{\delta \ln f_{\vec{X}}(\vec{X}, \theta)}{\delta \theta} = k(\theta) \cdot (\hat{\theta}(\vec{X}) - \theta)}_{\textit{критерий сущ-ния эфф. по Рао-Крамеру}} \tag{21}$$

При выполнении этого условия оценка $\hat{\theta}$ является эффективной (по Рао-Крамеру) и $k(\theta) = \frac{1}{D(\theta)} = I(\theta)$. Покажем, что $k(\theta) = \frac{1}{D(\theta)}$: Из 21:

$$[\frac{\delta \ln f_{\vec{X}}(\vec{X}, \theta)}{\delta \theta}]^2 = k^2(\theta) \cdot (\hat{\theta}(\vec{X}) - \theta)^2 \implies$$

$$\implies \underbrace{M[\frac{\delta \ln f_{\vec{X}}(\vec{X}, \theta)}{\delta \theta}]^2}_{I(\theta)} = k^2(\theta) \underbrace{M\{[\hat{\theta} - \theta]^2\}}_{D(\hat{\theta})} \implies$$

$$\implies I(\theta) = k^2(\theta) D(\hat{\theta}) \implies \langle \textit{ т. к. } \hat{\theta} \textit{ эффективная по Р.-К., то } I(\theta) = \frac{1}{D\hat{\theta}} \rangle \implies$$

$$\implies k^2(\theta) = \frac{1}{(D(\hat{\theta}))^2} \implies k(\theta) = \frac{1}{\underbrace{D(\hat{\theta})}_{I(\theta)}}$$

Замечание 4. *Если X — дискретная случайная величина, θ — неизвестный параметр закона распределения случайной величины X, то для всевозможных несмещённых оценок $\hat{\theta}$ параметра θ также будет выполняться неравенство Рао-Крамера, где под $I(\theta)$ понимается*

$$I(\theta) = M\{\left(\frac{\delta \ln(P\{X_1 = t_1\} \cdot \ldots \cdot P\{X_n = t_n\})}{\delta \theta}\right)^2\}$$

Лекция №23, 25.03.2019

3.3.5 Методы построения точечных оценок

Мы изучим два метода:

1. Метод моментов;
2. Метод максимального правдоподобия.

3.3.5.1 Метод моментов

Пусть

1. X — случайная величина, закон распределения которой известен с точностью до вектора $\vec{\theta} = (\theta_1, \ldots, \theta_r)$ неизвестных параметров.

   (требуется оценить значения параметров $\theta_1, \ldots, \theta_r$).

2. У случайной величины X $\exists r$ первых моментов.

Для построения точечных оценок параметров $\theta_1, \ldots, \theta_r$ с использованием метода моментов необходимо сделать следующее:

1. Найти выражения для r первых теоретических моментов случайной величины X (т. к. функция распределения случайной величины X зависит от параметров $\theta_1, \ldots, \theta_r$, то и теоретические моменты также будут зависеть от этих параметров):

   $m_1(\theta_1, \ldots, \theta_r) = M[X]$

   $\ldots$

   $m_r(\theta_1, \ldots, \theta_r) = M[X^r]$

2. Нужно приравнять выражения для теоретических моментов и их выборочных аналогов:

$$\begin{cases} m_1(\theta_1, \ldots, \theta_r) = \hat{m}_1(\vec{X}) \\ \ldots \\ m_r(\theta_1, \ldots, \theta_r) = \hat{m}_r(\vec{X}) \end{cases} \tag{22}$$

   Это — система из r нелинейных уравнений относительно r неизвестных $\theta_1, \ldots, \theta_r$.

3. Решаем получившуюся систему относительно $\theta_1, \ldots, \theta_r$:

$$\begin{cases} \theta_1 = \hat{\theta}_1(\vec{X}) \\ \ldots \\ \theta_r = \hat{\theta}_r(\vec{X}) \end{cases}$$

Замечание 1. *Некоторые уравнения системы 22 удобно записывать не относительно начальных моментов, а относительно центральных:*

$$\mathring{m}_j(\theta_1, \ldots, \theta_r) = \hat{\nu}_j(\vec{X})$$

где $\mathring{m}_j = M[(X - MX)^j]$, $\hat{\nu}_j(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^j$. *Выбор между начальными моментами для записи уравнения делают исходя из того, что система 22 должна решаться как можно проще.*

Замечание 2. *Поэтому выборочные моменты* $\hat{m}_k$ *и* $\hat{\nu}_k$ *являются состоятельными оценками своих теоретических аналогов, то и оценки параметров, полученные с использованием метода моментов, также будут состоятельными (при условии, что решение системы 22 непрерывно зависит от правых частей).*

Замечание 3. *Поскольку оценки* $\hat{m}_k$, $\hat{\nu}_k$ *при* $k \geqslant 2$ *являются смещёнными оценками соответствующих моментов, то и оценки параметров, полученных с использованием метода моментов, также могут быть смещёнными.*

Пример. $X \sim R[a, b]$ *С использованием метода моментов построить точечные оценки для* a *и* b.

Решение.

1.

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

a, b — *неизвестные параметры* $\implies r = 2$. *Нужно составить систему из двух уравнения для первого и второго моментов:*

$m_1(a, b) = MX = \frac{a+b}{2}$ — *первый начальный момент;*

$\mathring{m}_2(a, b) = DX = \frac{(b-a)^2}{12}$ — *второй центральный момент;*

2. *Составим систему*

$$\begin{cases} \frac{a+b}{2} = \hat{m}_1(\vec{X}) \\ \frac{(b-a)^2}{12} = \hat{\nu}_2(\vec{X}) \end{cases}$$

$$(\hat{m}_1(\vec{X}) = \overline{X},\ \hat{\nu}_2(\vec{X}) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}_{\hat{\sigma}^2(\vec{X})})$$

3. *Решаем получившуюся систему относительно a и b:*

$$\begin{cases} a + b = 2\overline{X} \\ (b-a)^2 = 12\sigma^2 \end{cases}$$

Приходим к

$$\begin{cases} a = \overline{X} \mp \sqrt{3}\hat{\sigma} \\ b = \overline{X} \pm \sqrt{3}\hat{\sigma} \end{cases}$$

Т. к. $a < b$, то

$$\begin{cases} \hat{a}(\vec{X}) = \overline{X} - \sqrt{3}\hat{\sigma}(\vec{X}) \\ \hat{b}(\vec{X}) = \overline{X} + \sqrt{3}\hat{\sigma}(\vec{X}) \end{cases}$$

3.3.5.2 Метод максимального правдоподобия

Пусть

1. X — случайная величина, закон распределения которой зависит от неизвестных параметров $\hat{\theta} = (\theta_1, \ldots, \theta_r)$;
2. $\vec{X} = (X_1, \ldots, X_n)$ — случайная выборка из генеральной совокупности X.

Определение. *Функцией правдоподобия случайной выборки $\vec{X}$ называется функция*

$$L(\vec{X}, \theta_1, \ldots, \theta_r) = p(X_1, \vec{\theta}) \cdot \ldots \cdot p(X_n, \vec{\theta})$$

где $p(x, \vec{\theta}) = P\{X = x\}$, если X — дискретная случайная величина; $p(x, \vec{\theta}) = f(x, \vec{\theta})$, если X — непрерывная случайная величина (f — функция плотности распределения случайной величины X).

Замечание. *Очевидно, что если в функцию L подставить удачное значение вектора $\vec{\theta}$ (т. е. подставить значения параметров $\theta_1, \ldots, \theta_n$, которые близки к их теоретическим значениям), то значение функции L будет относительно большим. Если же использовать «неудачные» значения вектора $\vec{\theta}$, то соответствующие значение L будет сравнительно малым.*

В методе максимального правдоподобия в качестве точечной оценки вектора $\vec{\theta}$ неизвестных параметров используют такое значение, которое максимизирует функцию правдоподобия L, т. е.

$$\forall \vec{x} \in X_n \; L(\vec{X}, \hat{\vec{\theta}}) \geqslant L(\vec{X}, \vec{\theta}), \; \vec{\theta} \in \Theta$$

Для построения такой оценки необходимо решить задачу

$$L(\vec{X}, \vec{\theta}) \to \max_{\vec{\theta} \in \Theta}$$

Эта задача эквивалентна задаче

$$\ln L(\vec{X},\, \vec{\theta}) \to \max_{\vec{\theta}\in\Theta}$$

которая часто решается проще.

В некоторых случаях для нахождения оптимального значения вектора $\vec{\theta}$ можно использовать необходимые условия экстремума ФНП?:

$$\frac{\delta L(\vec{X},\, \vec{\theta})}{\delta\theta} = 0,\ j = \overline{1;\, r} \tag{23}$$

или (для эквивалентной задачи)

$$\frac{\delta \ln L(\vec{X},\, \vec{\theta})}{\delta\theta_j} = 0,\ j = \overline{1;\, r} \tag{24}$$

Замечание. *Уравнения 23 (или уравнения 24) называют уравнения правдоподобия.*

Пример. $X \sim R[a,\, b]$. *Построить для a, b оценки максимального правдоподобия.*

Решение.

1.

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

2. Запишем функцию правдоподобия

$$L(\vec{X},\vec{\theta}) = f(X_1,\, a,\, b)\cdot\ldots\cdot f(X_n, a, b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leqslant X_{(1)},\, X_{(n)} \leqslant b \\ 0 & \textit{иначе} \end{cases}$$

$$\ln L = \begin{cases} -n\ln(b-a), & a \leqslant X_{(1)},\, X_{(n)} \leqslant B \\ -\infty & \textit{иначе} \end{cases}$$

Составим уравнения правдоподобия:

$$\begin{cases} \frac{\delta \ln L}{\delta a} = \frac{n}{n-a} = 0 \\ \frac{\delta \ln L}{\delta b} = \frac{-n}{b-a} = 0 \end{cases}$$

$\implies b - a = \infty$? *Поскольку от неизвестных параметров a, b зависят границы области, в которой* $f(x, a, b) > 0$*, то дифференцирование по этим параметрам затруднительно. По этой причине будем максимизировать функцию L «в лоб», не используя условия экстремума.*

3.

$$L(\vec{X}, a, b) = \begin{cases} \frac{1}{(b-a)^n}, & x \in [a;\, b] \\ 0 & \textit{иначе} \end{cases}$$

(a) *Если $X_{(1)} < a$ или $X_{(n)} > b$, то $L = 0$. По этой причине $a \leqslant X_{(1)}$, $X_{(n)} \leqslant b$;*

(b) *Чем меньше размер отрезка $[a,\, b]$, т. е. чем меньше $b - a$, тем больше $\frac{1}{b-a}$, следовательно, тем больше $L(x,\, a,\, b) = \frac{1}{(b-a)^n}$. Таким образом, для максимизации функции L нужно брать a, b так, чтобы $b - a \to \min$;*

(c) *Из а, б $\implies$ $\hat{a}(\vec{X}) = X_{(1)}$, $\hat{b}(\vec{X}) = X_{(n)}$.*

3.4 Интервальные оценки

3.4.1 Основные понятия

Задача. ***Вторая задача математической статистики.***

Пусть X — случайная величина, закон которой известен с точностью до вектора $\vec{\theta} = (\theta_1,\, \ldots,\, \theta_r)$ неизвестных параметров (т. е. известен общий вид закона распределения случайной величины X, но этот закон создаёт неизвестные параметры $\theta_1,\, \ldots,\, \theta_r$). Требуется: оценить значение $\vec{\theta}$.

Ранее для решения этой задачи использовались точечные оценки, недостаток которых состоит в том, что они не дают вероятностных характеристик точности оценивания параметров. Помимо точечных оценок для решения второй задачи математической статистики используется другой подход, занимающийся в построении доверительных интервалов.

Для простоты будем считать, что $r = 1$, т. е. $\vec{\theta} = (\theta)$.

Пусть $\gamma \in (0; 1)$. ([45])

Определение. *γ-доверительным интервалом (доверительным интервалом уровня γ) для параметра θ называется пара статистик*

$$\underline{\theta}(\vec{X}),\, \overline{\theta}(\vec{X})$$

таких, что

$$P\{\theta \in (\underline{\theta}(\vec{X}),\, \overline{\theta}(\vec{X}))\} = \gamma$$

Замечание 1. *Таким образом, γ-доверительный интервал — это интервал, который покрывает теоретическое значение параметра θ с вероятностью γ.*

Замечание 2. *Границы $\underline{\theta}(\vec{X})$, $\overline{\theta}(\vec{X})$ γ-доверительного интервала являются функциями случайной выборки, т. е. сами являются случайными величинами.*

[45] γ — строчная буква «гамма» греческого алфавита. — Прим. ред.

Замечание 3. *Вероятностной характеристикой γ-доверительного интервала является случайная величина $l(\vec{X}) = \overline{\theta}(\vec{X}) - \underline{\theta}(\vec{X})$, которая называется размахом этого интервала.*

Замечание 4. *Вероятность совершить ошибку при γ-доверительном интервале:*

$$P\{\theta \notin (\underline{\theta}(\vec{X}),\, \overline{\theta}(\vec{X}))\} = 1 - \gamma$$

Замечание 5. *Иногда удобно строить так называемые односторонние доверительные интервалы.*

Определение. *Односторонним нижним (верхним) γ-доверительным интервалом для параметра θ называется статистика $\underline{\theta}(\vec{X})$ (соответственно $\overline{\theta}(\vec{X})$) такая, что $P\{\theta \geqslant \underline{\theta}(\vec{X})\} = \gamma$ (соответственно $P\{\theta \leqslant \overline{\theta}(\vec{X})\} = \gamma$).*

3.4.2 Построение доверительных интервалов

Пусть

1. θ — неизвестный параметр закона распределения случайной величины X;
2. $g(\vec{X},\, \theta)$ — некоторая статистика.

Определение. *Статистику $g(\vec{X},\, \theta)$ будем называть центральной, если закон её распределения не зависит от θ, т. е. $F_g(x,\, \theta) \equiv F_g(x)$, где F_g — функция распределения случайной величины g.*

Пример. *Пусть $X \sim N(m,\, \sigma^2)$, где m — неизвестно, σ^2 — известно. Рассмотрим статистику $g(\vec{X},\, m) = \frac{m-\overline{X}}{\sigma} \cdot \sqrt{n}$, где $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Найдём закон распределения случайной величины g.*

1. *$g(\vec{X},\, m) = \frac{m\sqrt{n}}{\sigma} - \frac{1}{\sigma\sqrt{n}}(X_1 + \ldots + X_n)$, т. е. g является линейным преобразованием нормальной случайных величин $X_1,\, \ldots,\, X_n$, т. е., следовательно, g сама является нормальной случайной величиной ($g \sim N(m_g,\, \sigma_g^2)$).*

2. $m_g = M[g] = M[\frac{m-\overline{X}}{\sigma}\sqrt{n}] = \frac{m-\cancelto{m}{M[\overline{X}]}}{\sigma}\sqrt{n} = 0$

   $\sigma_g^2 = D[g] = D[\frac{m-\overline{X}}{\sigma}\sqrt{n}] = (\frac{\sqrt{n}}{\sigma})^2 D[\cancel{m} - \overline{X}] = \frac{n}{\sigma^2}\underbrace{D[\overline{X}]}_{\frac{\sigma^2}{n}} = 1$

3. *Таким образом, $g(\vec{X},\, m) \sim N(0,\, 1)$, не зависит от неизвестного параметра m.*

   Таким образом, g — центральная статистика.

Лекция №24, 01.04.2019

Пусть

1. $g(\vec{X},\,\theta)$ является центральной;

2. Случайная величина g является непрерывной случайной величиной, т. е. $F_g(x)$ — непрерывная (из этого следует, что уравнение $F_g(x) = \varpi$ ([46]) имеет единственное решение $\forall\varpi \in (0;\,1)$; q_ϖ — решение этого уравнение — квантиль уровня ϖ);

3. $q(\vec{X},\,\theta)$ как функция параметра θ является монотонно возрастающей;

4. Выбраны значения $\alpha_1,\,\alpha_2 \geqslant 0$ такие, что $\alpha_1 + \alpha_2 = 1 - \gamma$.

Тогда из свойств непрерывных случайных величин

$$\gamma = P\{q_{\alpha_1} < g(\vec{X},\,\theta) < q_{1-\alpha_2}\} = \langle q \text{ может возрастать с ростом } \theta\rangle =$$
$$= P\{\underbrace{g^{-1}(\vec{X},\,q_{\alpha_1})}_{\underline{\theta}(\vec{X})} < \theta < \underbrace{g^{-1}(\vec{X},\,q_{1-\alpha_2})}_{\overline{\theta}(\vec{X})}\}$$

Можно заметить, что согласно определению γ-доверительного интервала в качестве его границ можно взять отрезки

$$\underline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_{\alpha_1}),\ \overline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_{1-\alpha_2})$$

Замечание 1. *Можно заметить, что построенный γ-доверительный интервал зависит от выбора α_1 и α_2. Как правило, используют значения $\alpha_1 = \alpha_2$; тогда $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, где $\alpha = 1 - \gamma$.*

Замечание 2. *Если взять $\alpha_2 = 0$, то $\alpha_1 = 1 - \gamma$. В этом случае $q_{1-\alpha_2} = +\infty$ и доверительный интервал имеет вид $(\underline{\theta}(\vec{X}),\,+\infty)$, где $\underline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_{1-\gamma})$ — нижняя граница γ-доверительного множества?.*

Замечание 3. *Если взять $\alpha_1 = 0$, то $\alpha_2 = 1 - \gamma$. В этом случае $q_{\alpha_1} = -\infty$ и доверительный интервал имеет вид $(-\infty,\,\overline{\theta}(\vec{X}))$, где $\overline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_\gamma)$ — верхняя граница γ-доверительного множества?.*

Пример. *Пусть $X \sim N(m,\,\sigma^2)$, где m — неизвестно, σ^2 — известно. Выше было показано, что статистика*

$$g(\vec{X},\,m) = \frac{m - \vec{X}}{\sigma}\sqrt{n} \sim N(0,\,1)$$

Построим с её помощью γ-доверительный интервал для m.

$$\gamma = P\{u_{\alpha_1} < g(\vec{X},\,m) < u_{1-\alpha_2}\} = P\{u_{\alpha_1} < \frac{m - \overline{X}}{\sigma}\sqrt{n} < u_{1-\alpha_2}\} =$$
$$= P\{\underbrace{\overline{X} + \frac{\delta u_{\alpha_1}}{\sqrt{n}}}_{\underline{m}(\vec{X})} < m < \underbrace{\overline{X} + \frac{\delta u_{1-\alpha_2}}{\sqrt{n}}}_{\overline{m}(\vec{X})}\}$$

[46] ϖ — буква «пи» греческого алфавита. — Прим. ред.

Таким образом,

$$\underline{m}(\vec{X}) = \overline{X} + \frac{\delta u_{\alpha_1}}{\sqrt{n}},\ \overline{m}(\vec{X}) = \overline{X} + \frac{\delta u_{1-\alpha_2}}{\sqrt{n}}$$

Замечание 1. *Размах γ-доверительного интервала, построенного в примере:*

$$l(\vec{X}) = \overline{m} - \underline{m} = \frac{\sigma}{\sqrt{n}}(u_{1-\alpha_2} - u_{\alpha_1})$$

Можно заметить, что в этом примере размах является детерминизированной величиной (т. е. величиной, которая не является случайной). При этом очевидно, что размах будет минимальным, если $\alpha_1 = \alpha_2$. В этом случае

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = \frac{1-\gamma}{2}$$

$$u_{\alpha_1} = u_{\frac{\alpha}{2}} = \langle \textit{в силу симметрии графика } f_g(x) \rangle = -u_{1-\frac{\alpha}{2}}$$

$$u_{\alpha_2} = u_{\frac{1+\gamma}{2}}$$

Тогда

$$l = \frac{2\sigma}{\sqrt{n}} \cdot u_{\frac{1+\gamma}{2}}$$

При этом

$$\underline{m}(\vec{X}) = \overline{X} - \frac{\sigma u_{\frac{1+\gamma}{2}}}{\sqrt{n}},\ \overline{m}(\vec{X}) = \overline{X} + \frac{\sigma u_{\frac{1+\gamma}{2}}}{\sqrt{n}}$$

Замечание 2. *Зависимость l от параметров. При $\gamma \to 1-0$ $l \to +\infty$ (видно из функции, т. к. $u_{\frac{1+\gamma}{2}} \to +\infty$). Это означает, что чем более достоверную оценку мы хотим получить, тем меньше определённой она будет — ситуация, типичная для математической статистики. Для повышения надёжности результатов при сохранении их детерминизированности следует увеличивать объём наблюдений n.*

Пример. *$X \sim N(m, \sigma^2)$, где m — неизвестно, σ^2 — неизвестна. Построить доверительный интервал для m.*

Использовать статистику

$$g(\vec{X}, m) = \frac{m - \overline{X}}{\sigma}\sqrt{n}$$

не получится, т. к. σ — неизвестна. Используем вместо σ исправленную оценку для среднеквадратичного отклонения:

$$S(\vec{X}) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$g(\vec{X}, m) = \frac{m - \overline{X}}{S(\vec{X})}\sqrt{n}$$

Запишем g в виде

$$g(\vec{X},\, m) = \frac{\overbrace{\frac{m-\overline{X}}{\sigma}\sqrt{n}}^{=\xi}}{\sqrt{\underbrace{\frac{(n-1)S^2(\vec{X})}{\sigma^2}}_{=\eta}}}\sqrt{n-1} = \left\langle \xi = \frac{m-\overline{X}}{\sigma}\sqrt{n},\ \eta = \frac{(n-1)S^2(\vec{X})}{\sigma^2} \right\rangle =$$

$$= \frac{\xi}{\sqrt{\eta}}\sqrt{n-1}$$

1. $\xi \sim N(0,\, 1)$ *(см. предыдущие примеры);*
2. *Можно показать, что* $\eta \sim \xi^2(n-1)$*, причём случайные величины* ξ *и* η *— независимые;*
3. *Таким образом, случайная величина* g *имеет распределение* $St(n-1)$*. Построим* γ*-доверительный интервал для* m*:* $\alpha_1 = \alpha_2 = \frac{1-\gamma}{2}$*. Из свойств непрерывных случайных величин:*

$$\gamma = P\{-t_{\frac{1+\gamma}{2}} < g(\vec{X},\, m) < t_{\frac{1+\gamma}{2}}\} = P\{-t_{\frac{1+\gamma}{2}} < \frac{m-\overline{X}}{S(\vec{X})}\sqrt{n} < t_{\frac{1+\gamma}{2}}\} =$$

$$P\{\underbrace{\overline{X} - \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}}}_{\underline{m}(\vec{X})} < m < \underbrace{\overline{X} + \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}}}_{\overline{m}(\vec{X})}\}$$

Таким образом,

$$\underline{m}(\vec{X}) = \overline{X} - \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}},\ \overline{m}(\vec{X}) = \overline{X} + \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}}$$

Пример. $X \sim (m,\, \sigma^2)$*, где* m *— неизвестно,* σ^2 *— неизвестно. Оценить* σ^2*. Рассмотрим статистику*

$$g(\vec{X},\, \sigma^2) = \frac{S^2(\vec{X})}{\sigma^2}\cdot(n-1) \sim \xi^2(n-1)$$

(см. предыдущий пример). Покажем $\alpha_1 = \alpha_2 = \frac{\alpha}{2} \implies \alpha_1 = \frac{1-\gamma}{2}$*.*

$$\gamma = P\{h_{\frac{1-\gamma}{2}} < g(\vec{X},\, \sigma^2) < h_{\frac{1+\gamma}{2}}\} = P\{h_{\frac{1-\gamma}{2}} < \frac{S^2(n-1)}{\sigma^2} < h_{\frac{1+\gamma}{2}}\} =$$

$$= P\{\frac{1}{h_{\frac{1+\gamma}{2}}} < \frac{\sigma^2}{S^2(n-1)} < \frac{1}{h_{\frac{1+\gamma}{2}}}\} = P\{\underbrace{\frac{S^2(n-1)}{h_{\frac{1+\gamma}{2}}}}_{\underline{\sigma^2}} < \sigma^2 < \underbrace{\frac{S^2(n-1)}{h_{\frac{1+\gamma}{2}}}}_{\overline{\sigma^2}}\}$$

Таким образом,

$$\underline{\sigma^2} = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1+\gamma}{2}}},\ \overline{\sigma^2} = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1-\gamma}{2}}}$$

3.5 Проверка параметрических гипотез

3.5.1 Основные понятия

Пусть X — случайная величина, закон распределения которой неизвестен (известен не полностью).

Определение. *Статистической гипотезой называется любое утверждение относительно закона распределения случайной величины X.*

Определение. *Статистическая гипотеза называется простой, если она однозначно определяет закон распределения случайной величины X (однозначно задаёт функцию распределения случайной величины X как функцию своего аргумента). В противном случае статистическая гипотеза называется сложной.*

Определение. *Статистическая гипотеза называется параметрической, если она является утверждением относительно значений неизвестных параметров известного закона распределения.*

Пример 1. *Пусть $X \sim N(m, \sigma^2)$, где m, σ^2 — неизвестны. Рассмотрим гипотезы:*

$H_1 = \{m = 1, \sigma^2 = 5\}$ *(параметрическая простая гипотеза)*

$H_2 = \{m > 1, \sigma^2 = 5\}$ *(параметрическая сложная гипотеза)*

$H_3 = \{m = 1\}$ *(параметрическая сложная гипотеза)*

Пример 2. X *— случайная величина.*

$H_1 = \{X \sim Exp(5)\}$ *(непараметрическая простая гипотеза)*

$H_1 = \{X$ *— нормальная сл-ная величина*$\}$ *(непараметрическая сложная гипотеза)*

$H_1 = \{X \sim \Pi(\lambda),\ \lambda > 5\}$ *(непараметрическая сложная гипотеза)*

Проверку статистических гипотез проводят следующим образом:

1. Формулируют основную гипотезу H_0;

2. Формулируют конкурирующую (альтернативную) гипотезу H_1; $H_0 \cap H_1 = \emptyset$, но, возможно, $H_0 + H_1$ не исчерпывают все возможные случаи;

3. На основании имеющейся выборки $\vec{X} \in X_n$ принимают решение об истинности H_0 или H_1.

Определение. *Правило, посредством которого принимается решение об истинности H_0 или H_1, называется статистическим критерием проверки гипотез.*

Задают критерий проверки гипотез обычно с использованием так называемого критического множества $W \subseteq X_n$. При этом решающее правило имеет вид:

$$\vec{x} \in W \implies \begin{cases} \text{отклоняют } H_0 \\ \text{принимают } H_1 \end{cases}, \ \vec{x} \notin W \iff \vec{x} \in \underbrace{\overline{W}}_{X_n \setminus W} \implies \begin{cases} \text{принимают } H_0 \\ \text{отклоняют } H_1 \end{cases}$$

Замечание 1. *Задать критерий проверки гипотезы и задать критическое множество — одно и то же.*

Замечание 2. *При использовании любого критерия возможны ошибки двух видов:*

1. *Принять конкурирующую гипотезу при использовании основной гипотезы — ошибка I рода. Вероятность совершения этой ошибки обозначим $P\{\vec{x} \in W | H_0\} = \alpha$;*

2. *Принять H_0 при истинности H_1 — ошибка II рода. Вероятность её совершения: $P\{\vec{x} \notin W | H_1\} = P\{\vec{x} \in X_n \setminus W | H_1\} = \beta$.*

Замечание 1. *Множество $X_n \setminus W$ называется доверительным множеством;*

Замечание 2. *Конечно, при построении критерия хотелось мы сделать так, чтобы α и β минимизировались, однако это невозможно. По этой причине при построении критерия исходят из принципа*

$$\begin{cases} \beta \to \min \\ \alpha = const \end{cases}$$

или (что эквивалентно)

$$\begin{cases} \beta = const \\ \alpha \to \min \end{cases}$$

При этом α называют уровнем значительного критерия, а $1-\beta$ называют мощностью критерия (таким образом, при построении критерия максимизируют его мощность при фиксированном уровне значимости).

Лекция №25, 08.04.2019

Разрыва материала между 24 и 26 лекциями не видно. Эта лекция вообще была?

Лекция №26, 15.04.2019

3.5.2 Критерий Неймана-Пирсона проверки двух простых гипотез

Пусть

1. X — случайная величина;
2. $F(x,\,\theta)$ — функция распределения случайной величины X (известен общий вид функции F, но она зависит от неизвестного параметра θ).

Рассмотрим две простые параметрические гипотезы:

$$H_0 = \{\theta = \theta_0\} \text{ и } H_1 = \{\theta = \theta_1\}$$

где $\theta_0 \neq \theta_1$. Введём в рассмотрение статистику

$$\varphi(\vec{X}) = \frac{L(\vec{X},\,\theta_1)}{L(\vec{X},\,\theta_0)}$$

где $L(\vec{X},\,\theta)$ — функция правдоподобия.

Определение. *Статистика φ называется отношением правдоподобия.*

Если значения функции достаточно большие, то это ассоциируется с истинностью H_1. Очевидно, что «большие» значения статистики φ ассоциируются с истинностью конкурирующей гипотезы H_1, поэтому критическое множество должно иметь вид:

$$W = \{\vec{x} \in X_n \colon \varphi(\vec{x}) \geqslant C_\varphi\}$$

где константа C_φ выбирается из условия

$$\alpha = P\{\varphi(\vec{X}) \geqslant C_\varphi | \theta = \theta_0\} \tag{25}$$

Замечание. *При построении критерия зафиксировано некоторое $\alpha = const$ — вероятность совершения ошибки I рода. Чтобы построенный критерий имел уровень значимости α, необходимо, чтобы $P\{\vec{X} \in W | H_0\} = \alpha$ — в общем случае. В рассматриваемом случае $\vec{X} \in W \iff \varphi(\vec{X}) \geqslant C_\varphi$.*

$$H_0 \text{ — истина} \iff \theta = \theta_0$$

поэтому

$$P\{\vec{X} \in W | H_0\} = P\{\varphi(\vec{X}) \geqslant C_\varphi | \theta = \theta_0\}$$

Замечание 1. *Построенный критерий называется критерием Неймана-Пирсона.*

Замечание 2. *Если X — непрерывная случайная величина, то плотность распределения случайной выборки $\vec{X}$ имеет вид*

$$f_{\vec{X}}(x_1,\,\ldots,\,x_n,\,\theta) = \overbrace{f(x_1,\,\theta) \cdot \ldots \cdot f(x_n,\,\theta)}^{=L(x_1,\,\ldots,\,x_n,\,\theta)}$$

где $f(x,\,\theta)$ —- функция плотности распределения случайной величины X.

По этой причине условие 25 может быть записано в следующем виде:

$$\int\limits_{\varphi(t_1,\,\dots,\,t_n)\geqslant C_\varphi}\!\!\dots\!\!\int L(t_1,\,\dots,\,t_n,\,\theta)\,dt_1\dots dt_n = 1$$

Пример. *Пусть $X \sim N(m,\,\sigma^2)$, где m — неизвестно, σ^2 — известно. Рассмотрим задачу проверки двух простых гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m = m_1\}$$

где $m_0 < m_1$. В этом примере функция правдоподобия имеет вид:

$$L(x_1,\,\dots,\,x_n,\,m) = [\frac{1}{\sqrt{2\pi}\sigma}]^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-m)^2}$$

Тогда отношение правдоподобия:

$$\varphi(\vec{X}) = \frac{L(\vec{X},\,m_1)}{L(\vec{X},\,m_0)} = \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-m_1)^2}}{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-m_0)^2}} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(X_i-m_1)^2-(X_i-m_0)^2]} =$$
$$= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(X_i-m_1)^2-(X_i-m_0)^2]} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [X_i^2-2X_im_1+m_1^2-X_i^2+2X_im_0-m_0^2]} =$$
$$= e^{\frac{m_1-m_0}{\sigma^2}(\sum_{i=1}^n X_i)-\frac{n}{2\sigma^2}[m_1^2-m_0^2]} \tag{26}$$

Выше было показано, что критическое множество должно иметь вид

$$W = \{\vec{x}\colon \varphi(\vec{x}) \geqslant C_\varphi\}$$

где $C_\varphi = const$ выбирается из условия

$$P\{\varphi\vec{X} \geqslant C_\varphi | H_0\} = \alpha$$

Условие

$$\varphi(\vec{x}) \geqslant C_\varphi \iff \ln\varphi(\vec{X}) \geqslant \ln C_\varphi \iff \langle \text{см. 26}\rangle \iff$$
$$\iff \ln[e^{\frac{m_1-m_0}{\sigma^2}(\sum_{i=1}^n X_i)-\frac{n}{2\sigma^2}[m_1^2-m_0^2]}] \geqslant \ln C_\varphi \iff$$
$$\iff \frac{m_1-m_0}{\sigma^2}\left(\sum_{i=1}^n X_i\right) - \frac{n}{2\sigma^2}[m_1^2-m_0^2] \geqslant \ln C_\varphi \iff$$
$$\iff \frac{m_1-m_0}{\sigma^2}\left(\sum_{i=1}^n X_i\right) \geqslant \ln C_\varphi + \frac{n}{2\sigma^2}[m_1^2-m_0^2] \iff$$
$$\iff \langle m_1 > m_0 \implies m1 - m_0 > 0\rangle \iff \sum_{i=1}^n X_i \geqslant \underbrace{\frac{\sigma^2}{m_1-m_0}[\ln C_\varphi - \frac{n}{2\sigma^2}[m_1^2-m_0^2]]}_{c=const}$$

Таким образом,

$$W = \{\vec{x} \in X_n \colon \sum_{i=1}^{n} x_i \geqslant c\}$$

где c выбирается из условия

$$\alpha = P\{\varphi(\vec{X}) \geqslant C_\varphi | H_0\} = P\{\sum_{i=1}^{n} X_i \geqslant C | m = m_0\}$$

(эти события эквивалентны). Если истина H_0, т. е. $m = m_0$, то случайная величина

$$\sum_{i=1}^{n} X_i \sim (nm_0, n\sigma^2)$$

Таким образом,

$$\alpha = P\{\sum_{i=1}^{n} X_i \geqslant C | m = m_0\} = 1 - P\{\sum_{i=1}^{n} X_i < C | m = m_0\} = 1 - \Phi(\frac{c - nm_0}{\sqrt{n\sigma^2}})$$

то есть

$$\Phi(\frac{c - nm_0}{\sigma\sqrt{n}}) = 1 - \alpha$$

Таким образом

$$\frac{c - nm_0}{\sigma\sqrt{n}} = u_{1-\alpha}$$

($u_{1-\alpha}$ — квантиль уровня $1-\alpha$ стандартного нормального распределения). $c = \sigma u_{1-\alpha}\sqrt{n} + nm_0$

Таким образом, критерий имеет вид:

$$\sum_{i=1}^{n} X_i \geqslant \delta u_{1-\alpha}\sqrt{n} + nm_0 \implies \begin{cases} \textit{принять } H_1, \\ \textit{отклонить } H_0 \end{cases}$$
$$\sum_{i=1}^{n} X_i < \delta u_{1-\alpha}\sqrt{n} + nm_0 \implies \begin{cases} \textit{принять } H_0, \\ \textit{отклонить } H_1 \end{cases}$$

При этом вероятность совершения ошибки второго рода:

$$\beta = P\{\vec{X} \notin W | H_1\} = P\{\sum_{i=1}^{n} X_i < C | m = m_1\} =$$

$$= \langle c = \delta u_{1-\alpha}\sqrt{n} + nm_0; \textit{при } m = m_1 \colon \sum_{i=1}^{n} X_i \sim N(nm_1, n\sigma^2)\rangle =$$

$$= \Phi(\frac{(\sigma u_{1-\alpha}\sqrt{n} + nm_0) - nm_1}{\sigma\sqrt{n}}) = \Phi(\frac{\sigma u_{1-\alpha}\sqrt{n} - n(m_1 - m_0)}{\sigma\sqrt{n}}) = \Phi(u_{1-\alpha} - \sqrt{n}\frac{m_1 - m_0}{\sigma})$$

Замечание. *Если в условиях предыдущего примера $m_1 < m_0$, то критическое множество должно задаваться условием $\sum_{i=1}^n x_i \leqslant C$, где $C = const$ выбирается из условия $P\{\sum_{i=1}^n X_i > C|H_0\} = \alpha$, т. е. $P\{\sum_{i=1}^n X_i < C|m = m_0\} = \alpha$, т. е. $\Phi(c - \frac{nm_0}{\sigma\sqrt{n}}) = \alpha$, откуда*

$$\frac{C - nm_0}{\sigma\sqrt{n}} = u_\alpha \implies C = \underbrace{\delta u_\alpha + \sqrt{n}}_{-\delta u_{1-\alpha} - \sqrt{n}} + nm_0$$

(т. к. $u_\alpha = u_{1-\alpha}$ в случае ст. норм? распределения).

3.5.3 Проверка сложный параметрических гипотез

Пусть

1. X — случайная величина;

2. $F(x, \theta)$ — функция распределения случайной величины X (общий вид функции F известен, но F зависит от неизвестного параметра θ).

Рассмотрим задачу проверки двух сложных гипотез:

$$H_0 = \{\theta \in \Theta_0\} \text{ и } H_1 = \{\theta \in \Theta_1\}$$

где $\Theta_0 \cap \Theta_1 = \emptyset$.

Пример 1. $\Theta_0 = \{\theta > \theta_0\}$, $\Theta_1 = \{\theta \leqslant \theta_0\}$.

Пример 2. $\Theta_0 = \{\theta < \theta_0\}$, $\Theta_1 = \{\theta > \theta_1\}$, *где* $\theta_0 < \theta_1$.

В этом случае, критерий, как и раньше, задаётся с использованием критического множества W, а решающее правило имеет вид:

$$\vec{x} \in W \implies \begin{cases} \text{отклоняют } H_0 \\ \text{принимают } H_1 \end{cases}, \ \vec{x} \notin W \iff \vec{x} \in \underbrace{\overline{W}}_{X_n \backslash W} \implies \begin{cases} \text{принимают } H_0 \\ \text{отклоняют } H_1 \end{cases}$$

При этом ошибки первого и второго родов определяются так же, как и раньше, но теперь их вероятности зависят от θ:

$$\alpha(\theta) = P\{\vec{X} \in W|\theta \in \Theta_0\}$$
$$\beta(\theta) = P\{\vec{X} \in X_n \setminus W|\theta \in \Theta_1\}$$

Определение. *Величина*

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$$

называется размером критерия.

Определение. *Функция*

$$M(\theta) = P\{\vec{X} \in W|\theta\} \tag{27}$$

называется функцией мощности критерия.

Замечание 1. *Условие 27, принятое в математической статистике, было бы удачнее записать в виде*

$$M(t) = P\{\vec{X} \in W|\theta = t\}$$

т. е. $M(t)$ — вероятность события $\{\vec{X} \in W\}$ при условии, что неизвестный параметр имеет значение θ.

Замечание 2. *Через функцию мощности можно выразить вероятности совершения ошибок первого и второго рода:*

$$\alpha(\theta) = M(\theta),\ \theta \in \Theta_0$$
$$\beta(\theta) = 1 - M(\theta),\ \theta \in \Theta_1$$

Лекция №27, 22.04.2019

Определение. *Критерий, который при заданном размере α максимизирует функцию мощности одновременно по всем возможным критериям при всех $\theta \in \Theta_1$, называется равномерно наиболее мощным.*

Замечание 1. *Таким образом, равномерно наиболее мощный критерий, если он существует, при заданном размере α минимизирует вероятность совершения ошибки второго рода по всем возможным критериям при всех значениях $\theta \in \Theta_2$.*

Замечание 2. *Равномерно наиболее мощный критерий существуют лишь в некоторых частных случаях при проверке гипотез о значениях одномерных параметров.*

Пример. *Пусть $X \sim N(m, \sigma^2)$, где m — неизвестно, σ^2 — известно. Рассмотрим задачу проверки гипотез*

$$\underbrace{H_0 = \{m = m_0\}}_{\text{простая гипотеза}} \text{ и } \underbrace{H_1 = \{m > m_0\}}_{\text{сложная гипотеза}}$$

1. *Ранее была решена задача проверки двух простых гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m = m_1\}$$

где $m1 > m_0$. При этом критическое множество имело вид

$$W = \{\vec{x} \in X_n\colon \sum_{i=1}^{n} x_i \geqslant nm_0 + u_{1-\alpha}\sigma\sqrt{n}\} \tag{28}$$

2. *Т. к. построенное выше критическое множество не зависит от m_1, то фактически этот критерий является равномерно наиболее мощным для проверки гипотез $H_0 = \{m = m_0\}$ и $H_1 = \{m > m_0\}$. Таким образом, для рассматриваемой задачи критическое множество имеет вид 28.*

Замечание. *Если в условиях предыдущего примера рассматривается задача проверки гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m < m_0\}$$

то аналогичным образом приходим к выводу, что критическое множество будет иметь вид

$$W = \{\vec{x} \in X_n \colon \sum_{i=1}^{n} x_i \leqslant nm_0 - u_{1-\alpha}\sigma\sqrt{n}\}$$

(см. аналогичное замечание после примера о проверке $H_0 = \{m = m_0\}$ против $H_1 = \{m = m_1\}$ при $m_0 > m_1$).

Пример. *Пусть $X \sim N(m, \sigma^2)$, m — неизвестно, σ^2 — известно. Рассмотрим задачу проверки $H_0 = \{m = m_0\}$ против $H_1 = \{m \neq m_0\}$. В предыдущих примерах рассмотрение аналогичной задачи привело к критическим множествам вида:*

1. $H_1 = \{m > m_0\}$, $W = \{\vec{x} \in X_n \colon T(\vec{x}) \geqslant u_{1-\alpha}\}$, *где $u_{1-\alpha}$ — квантиль стандартного нормального распределения.*

   $$T(\vec{X}) = \frac{\overline{X} - m_0}{\sigma}\sqrt{n}$$

   (при истинности гипотезы H_0 $T(\vec{X}) \sim N(0, 1)$).

2. $H_1 = \{m < m_0\}$, $W = \{\vec{x} \in X_n \colon T(\vec{X}) \leqslant u_{1-\alpha}\}$, *где T и $u_{1-\alpha}$ имеют тот же смысл, что и в пункте а.*

Воспользуемся статистикой T и в рассматриваемом примере, когда $H_1 = \{m \neq m_0\}$. ?

Тогда при истинности конкурирующей гипотезы H_0 статистика T будет принимать большие по абсолютной величине значения, поэтому критическое множество можно записать в виде

$$W = \{\vec{x} \in X_n \colon |T(\vec{x})| \geqslant u_{1-\frac{\alpha}{2}}\}$$

В этом случае размер критерия (т. е. вероятность совершить ошибку первого рода) равен α.

Пример. *Пусть $X \sim N(m, \sigma^2)$, где m — неизвестно, σ^2 — известно. Рассмотрим задачу проверки*

$$\underbrace{H_0 = \{m = m_0\}}_{\text{сложная (т. к. дисперсия неизвестна)}} \quad \text{против} \quad \underbrace{H_1 = \{m > m_0\}}_{\text{сложная}}$$

В этом примере целесообразно воспользоваться статистикой

$$T(\vec{X}) = \frac{\overline{X} - m_0}{S(\vec{X})}\sqrt{n} \sim St(n-1) \text{ (при истинности } H_0\text{)}$$

Аналогично предыдущим примерам критическое множество можно задать в виде

$$W = \{\vec{x} \in X_n \colon T(\vec{X}) \geqslant t_{1-\alpha}^{(n-1)}\}$$

где $t_{1-\alpha}^{(n-1)}$ *— квантиль уровня* $1-\alpha$ *распределения* $St(n-1)$.

Замечание. *Пусть в условиях предыдущего примера требуется проверить следующие гипотезы:*

1. $H_0 = \{m = m_0\}$ *против* $H_1 = \{m < m_0\}$ *и*
2. $H_0 = \{m = m_0\}$ *против* $H_1 = \{m \neq m_0\}$.

Рассуждая аналогично, с использованием статистики

$$T(\vec{X}) = \frac{\overline{X} - m_0}{S(\vec{X})}\sqrt{n}$$

зададим критические множества в виде

1. $W = \{\vec{x} \in X_n \colon T(\vec{X}) \leqslant -t_{1-\alpha}^{(n-1)}\}$
2. $W = \{\vec{x} \in X_n \colon |T(\vec{X})| \geqslant t_{1-\frac{\alpha}{2}}^{(n-1)}\}$

Пример. *Пусть*

1. $X \sim N(m_1, \sigma_1^2)$;
2. $Y \sim N(m_2, \sigma_2^2)$;
3. m_1, m_2 *— неизвестно,* σ_1, σ_2 *— известны. Рассмотрим задачи проверки гипотез:*
   (a) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 > m2\}$;
   (b) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 < m_2\}$;
   (c) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 \neq m_2\}$.

   Рассмотрим случайную величину $Z = X - Y$, $MZ = MX - MY$, *поэтому сформулированные задачи эквивалентны задачам:*

   (a) $H_0 = \{m = 0\}$ *против* $H_1 = \{m > 0\}$;
   (b) $H_0 = \{m = 0\}$ *против* $H_1 = \{m < 0\}$;

(c) $H_0 = \{m = 0\}$ *против* $H_1 > \{n \neq 0\}$,

где $m = M[Z]$.

Рассмотрим статистику

$$T(\vec{X}, \vec{Y}) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$$

где n_1 *— объём выборки* $\vec{X}$, n_2 *— объём выборки* $\vec{Y}$. *Каков закон распределения случайной величины* G *при истинности* H_0*?* T *является линейной комбинацией нормальных случайных величин, следовательно,* T *сама имеет нормальное распределение.*

$$M[T] = \frac{1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}(M[\overline{X}] - M[\overline{Y}]) = \langle \text{при истинности } H_0\ m_1 = m_2 \rangle = 0$$

$$D[T] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot [D\overline{X} + D\overline{Y}] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}] = 1$$

Таким образом, при истинности H_0 *статистика* $T \sim N(0, 1)$. *По этой причине критические множества в каждой из рассматриваемых задач имеют вид:*

(a) $W = \{(\vec{x}, \vec{y}) \in X_n \colon T(\vec{x}, \vec{y}) \geqslant u_{1-\alpha}\}$, $m > 0$;

(b) $W = \{(\vec{x}, \vec{y}) \in X_n \colon T(\vec{x}, \vec{y}) \leqslant -u_{1-\alpha}\}$, $m < 0$;

(c) $W = \{(\vec{x}, \vec{y}) \in X_n \colon |T(\vec{x}, \vec{y})| \geqslant u_{1-\frac{\alpha}{2}}\}$, $m \neq 0$.

Лекция №28, 29.04.2019

Пример. *Пусть*

1. $X_1 \sim N(m_1, \sigma_1^2)$, *где* m_1, σ_1^2 *— неизвестные;*

2. $X^2 \sim N(m_2, \sigma_2^2)$, *где* m_2, σ_2^2 *— неизвестные.*

Рассмотрим задачи проверки гипотез

1. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 > m2\}$;

2. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 < m_2\}$;

3. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 \neq m_2\}$.

Рассуждая аналогично предыдущему примеру, заключаем, что статистика

$$\tilde{T}(\vec{X},\,\vec{Y}) = \underbrace{\frac{(n_1-1)S^2(\vec{X})}{\sigma_1^2}}_{\sim\chi^2(n_1-1)} + \underbrace{\frac{(n_2-1)S^2(\vec{Y})}{\sigma_2^2}}_{\sim\chi^2(n_2-1)} \sim \chi^2(n_1+n_2-2)$$

При истинности H_0 статистика

$$\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,\,1)$$

поэтому статистика

$$T(\vec{X},\,\vec{Y}) = \frac{(\overline{X}-\overline{Y})\sqrt{n_1+n_2-2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{(n_1-1)S^2(\vec{X})+(n_2-1)S^2(\vec{Y})}} \sim St(n_1+n_2-2)$$

Таким образом, критические множества для каждой из рассматриваемых задач выглядят так же, как и в предыдущем примере при условии $u \longleftrightarrow t^{(n_1+n_2-2)}$.

Пример. *Пусть*

1. $X \sim N(m_1,\,\sigma_1^2)$; $Y \sim N(m_2,\,\sigma_2^2)$;
2. $m_1,\,m_2,\,\sigma_1^2,\,\sigma_2^2$ — *неизвестны.*

Рассмотрим задачи проверки гипотез

1. $H_0 = \{\sigma_1 = \sigma_2\}$ *против* $H_1 = \{\sigma_1 > \sigma 2\}$;
2. $H_0 = \{\sigma_1 = \sigma_2\}$ *против* $H_1 = \{\sigma_1 < \sigma_2\}$;
3. $H_0 = \{\sigma_1 = \sigma_2\}$ *против* $H_1 = \{\sigma_1 \neq \sigma_2\}$.

Рассмотрим статистику

$$T(\vec{X},\,\vec{Y}) = \frac{S^2(\vec{X})}{S^2(\vec{Y})} \sim F(n_1-1,\,n_2-2) \text{ (распределение Фишера)}$$

Рассмотрим задачу а. Рассуждая аналогично предыдущим примерам, приходим к выводу, что критическое множество имеет вид

$$W = \{(\vec{x},\,\vec{y})\colon T(\vec{x},\,\vec{y}) \geqslant f_{1-\alpha}^{(n_1-1,\,n_2-1)}\}$$

В задаче б:

$$W = \{(\vec{x},\,\vec{y})\colon T(\vec{x},\,\vec{y}) \leqslant f_{\alpha}^{(n_1-1,\,n_2-1)}\}$$

В задаче в:

$$W = \{(\vec{x},\,\vec{y})\colon (T(\vec{x},\,\vec{y}) \leqslant f_{\frac{\alpha}{2}}^{(n_1-1,\,n_2-1)}) \vee (T(\vec{x},\,\vec{y}) \geqslant f_{1-\frac{\alpha}{2}}^{(n_1-1,\,n_2-1)})\}$$

3.6 Проверка непараметрических гипотез: критерии согласия

3.6.1 Основные понятия

До сих пор рассматривалась вторая задача математической статистики:

Дано X — случайная величина, закон распределения которой известен с точностью до вектора $\vec{\theta}$ неизвестных параметров. Требуется:оценить значение $\vec{\theta}$.

Рассмотрим первую задачу математической статистики:

Дано: X — случайная величина, закон распределения которой неизвестен. Требуется: найти закон распределения случайной величины X.

Решение этой задачи сводится к проверке основной гипотезы

$$H_0 = \{F(t) \equiv F_0(t)\} = \{(\forall t \in \mathbb{R})(F(t) = F_0(t))\}$$

где $F(t)$ — функция распределения случайной величины X, $F_0(t)$ — некоторая известная функция распределения, против конкурирующей гипотезы

$$H_1 = \neg H_0 = \{(\exists t \in \mathbb{R})(F(t) \neq F_0(t))\}$$

Замечание. *В таком виде задача формулирована для проверки простой гипотезы H_0. Однако на практике гораздо чаще возникает задача проверки гипотезы об общем виде функции распределения случайной величины X. Например,*

$$H_0 = \{X \text{ имеет нормальное распределение (с некоторыми значениями параметров)}\}$$

Или

$$H_0 = \{X \text{ имеет распределение Пуассона (для некоторого значения параметра)}\}$$

В этом случае гипотеза H_0 является сложной и (если $\vec{\theta}$ — вектора неизвестных параметров) имеет вид:

$$H_0 = \{(\exists \vec{\theta}_0)(\forall t \in \mathbb{R})(F(t) = F_0(t, \vec{\theta}_0)\}$$

где $F(t)$ — функция распределения случайной величины X, $F_0(t, \vec{\theta})$ — некоторая функция распределения, известная с точностью до вектора $\vec{\theta}_0$. При этом конкурирующая гипотеза:

$$H_1 = \neg H_0 = \{(\forall \vec{\theta})(\exists t \in \mathbb{R})(F(t) \neq F_0(t, \vec{\theta})\}$$

Проверка основной гипотезы H_0 сводится к оценке величины

$$\triangle(F_n, F_0)$$

рассогласования эмпирической функции распределения и предполагаемой функции распределения F_0.

Определение. *Криетрием согласия называется статистический критерий, предназначенный для проверки корректности гипотезы о том, что предполагаемый закон распределения $F_0(t, \vec{\theta})$ случайной величины X соответствует экспериментальным данным, представленным эмпирической функцией распределения $F_n(t)$.*

Замечание. *При выдвижении основной гипотезы H_0 могут быть использованы следующие соображения:*

1. *Анализ априорной информации об изученном процессе или явлении и её сопоставление и исходными предпосылками построения конкретных моделей (нормальной, биномиальной, экспоненциальной);*

2. *Построение эмпирической функции распределения по имеющейся выборке $\vec{x}$. На основании вида этой функции может быть выдвинута гипотеза H_0;*

3. *Построение гистограммы и использование её для выдвижения гипотезы H_0.*

3.6.2 Критерий Колмогорова для простой гипотезы

Пусть

1. X — непрерывная случайная величина;

2. $\vec{X}$ — случайная выборка из генеральной совокупности $\vec{X}$.

Рассмотрим задачу проверки гипотезы $H_0 = \{F(t) \equiv F_0(t)\}$ против $H_1 = \neg H_0$.

Замечание. *Здесь $F_0(t)$ — полностью известная функция распределения, которая не зависит ни от каких-то ни было неизвестных параметров. По этой причине H_0 — простая гипотеза.*

Для решения этой задачи рассмотрим статистику $\triangle(\vec{X})$, реализации которой определяются соотношением

$$\triangle(\vec{x}) = \sup |F_n(t) - F_0(t)|,\ t \in \mathbb{R}$$

где $F_n(t)$ — эмпирическая функция распределения, построенная по выборке $\vec{x}$. Очевидно, что «малые» значения статистики $\triangle$ свидетельствуют об истинности H_0, а «большие» — об истинности H_1. По этой причине критическое множество имеет вид

$$W = \{\vec{x} \in X_n \colon \triangle(\vec{x}) \geqslant \sigma_{1-\alpha}\}$$

где $\sigma_{1-\alpha}$ — квантиль уровня $1-\alpha$ закона распределения случайной величины $\triangle(\vec{X})$. При этом собственно решающее правило имеет стандартный вид:

$$\begin{aligned} \vec{x} \in W &\implies \text{принять } H_1,\ \text{отклонить} H_0 \\ \vec{x} \in \overline{W} &\implies \text{принять } H_0,\ \text{отклонить} H_1 \end{aligned}$$

Замечание. *О законе распределения случайной величины* $\triangle(\vec{X})$.

1. *Построенный критерий предполагает использование квантилей закона распределения случайной величины* $\triangle(\vec{X})$. *Вообще говоря, этот закон распределения зависит от:*

   (a) *От выбранной функции* $F_0(t)$;

   (b) *От теоретического закона распределения случайной величины* X.

2. *В математической статистике доказано утверждение:*

   Теорема. *Пусть*

   (a) $Y \sim R[0;\, 1]$;

   (b) $\hat{R}_n(t,\, \vec{Y})$ — *выборочная функция распределения случайной величины* Y, *отвечающая выборке* $\vec{Y}$.

   Тогда при исполнении H_0 *функция распределения случайной величины* $\triangle(\vec{X})$ *совпадает с функцией распределения случайной величины* $Z(n) = \sup\limits_{t \in [0;\, 1]} |\hat{R}_n(t,\, \vec{Y}) - t|$.

   Таким образом, для всевозможных законов распределения случайной величины X *и всевозможных функций распределения* F_0, *участвующих в построении основной гипотезы* H_0, *закон распределения случайной величины* $\triangle(\vec{X})$ *совпадает с законом распределения случайной величины* $Z(n)$.

3. *Для каждого фиксированного* $n \in \mathbb{N}$ *закон распределения случайной величины* $Z(n)$ *известен. Для* $n \leqslant 100$ *составлены таблицы значений соответствующей функции распределения.*

4. *Что делать, если* $n > 100$?

   (a) *Колмогоров доказал, что*

   $$\lim_{n \to \infty} P\{\sqrt{n} Z(n) < t\} = K(t),\ t > 0$$

   где $K(t) = \sum_{n=-\infty}^{+\infty} (-1)^k e^{-2k^2 t^2}$. *Другими словами, при* $n \to \infty$ *последовательность случайных величин* $\sqrt{n} Z(n)$ *слабо сходится к случайной величине* A, *функцией распределения которой является*

   $$F_a(t) = \begin{cases} 0, & t \leqslant 0 \\ k(t), & t > 0 \end{cases}$$

(b) *Таким образом, при достаточно больших n квантили случайной величины $\sqrt{n}Z(n)$ совпадают с квантилями соответствующих уровней случайной величины A. Если а м— квантиль случайной величины A, б — квантиль случайной величины $Z(n)$, то*

$$\sigma_{1-\alpha} = \frac{a_{1-\alpha}}{\sqrt{n}} \tag{29}$$

Таблица значений функции $K(t)$ составлены, их можно найти в литературе (интернете).

5. *Как показывает практика, соотношением 29 можно пользоваться при $n \geqslant 20$.*

Замечание. *Построенный критерий называется критерием Колмогорова для простой гипотезы.*

Лекция №29, 06.05.2019

3.6.3 Критерий хи-квадрат для простой гипотезы

Пусть

1. X — *дискретная* случайная величина;
2. X может принимать конечное множество значений $a_1, \ldots, a_l$ с неизвестными вероятностями $p_1, \ldots, p_l$.

Требуется проверить основную гипотезу

$$H_0 = \{p_1 = p_1^0, \ldots, p_l = p_l^0\}$$

(где $p_1^0, \ldots, p_l^0$ — некоторые известные значения) против конкурирующей гипотезы

$$H_1 = \neg H_0 = \{\exists k \in \{1, \ldots, l\} \colon p_k \neq p_k^0\}$$

Для решения этой задачи рассмотрим статистики $n_1(\vec{X}), \ldots, n_l(\vec{X})$, где выборочные значения $n_k(\vec{X})$

$$n_k(\vec{x}) = \{\text{кол-во элементов выборки } \vec{x} \text{ которые имеют значение } n_k\}$$

Замечание. *Очевидно, что $n_1(\vec{X}) + \ldots + n_l(\vec{X}) = n$, поэтому случайные величины $n_1(\vec{X}), \ldots, n_l(\vec{X})$ зависимы.*

Теорема. ***Теорема Пирсона.***

Пусть выполнены сделанные выше предположения. Тогда при истинности H_0 последовательность случайных величин

$$\sum_{k=1}^{l} \frac{(n_k(\vec{X}) - np_k)^2}{np_k}$$

слабо сходится к случайной величине, имеющей распределение $\chi^2(l-1)$.

Согласно этой теореме, при $n \to \infty$ случайная величина

$$\triangle(\vec{X}) = \sum_{k=1}^{l} \frac{(n_k(\vec{X}) - np_k^0)^2}{np_k^0} = n \sum_{k=1}^{l} \frac{(\frac{n_k(\vec{X})}{n} - p_k^0)^2}{p_k^0}$$

сходится к случайной величине, распределённой по закону $\chi^2(l-1)$.

Очевидно, что истинность гипотезы H_0 ассоциируется с «малыми» значениями статистики $\triangle(\vec{X})$, а истинность конкурирующей гипотезы H_1 — с «большими» положительными значениями. По этой причине критическое множество можно задать в виде

$$W = \{\vec{x} \in X_n \colon \triangle(\vec{x}) \geqslant h_{1-\alpha}^{(l-1)}\}$$

где $h_{1-\alpha}^{(l-1)}$ — квантиль уровня $1-\alpha$ распределения $\chi^2(l-1)$.

Замечание 1. *Построенный критерий называется критерием хи-квадрат (χ^2) или критерием Пирсона.*

Замечание 2. *В отличие от критерия Колмогорова, критерий хи-квадрат носит асимптотический характер, т. к. статистика $\triangle(\vec{X})$ имеет распределение, близкое к $\chi^2(l-1)$ лишь при достаточном больших n.*

Эмпирические рекомендации таковы:

1. *При $l \lesssim 20$: должно быть $np_k \geqslant 10$, $k = \overline{1; l}$ для использования критерия;*

2. *При $l > 20$, достаточно выполнения условия $np_k \geqslant 5$, $k = \overline{1; l}$.*

3.6.4 Критерий Колмогорова для сложной гипотезы

На практике задачи проверки простой гипотезы

$$H_0 = \{F(t) \equiv F_0(t)\}$$

встречается крайне редко; как правило, требуется проверить гипотезу о принадлежности закона распределения случайной величины X заданному классу (нормальному, экспоненциальному и т. д.) По этой причине основная гипотеза H_0 будет сложной:

$$H_0 = \{(\exists\vec{\theta})(\forall t \in \mathbb{R})(F(t) = F_0(t, \vec{\theta}))\}$$

где $F(t)$ — теоретический ("истинный") закон распределения случайной величины X, F_0 — предполагаемый закон распределения случайной величины, $\vec{\theta}$ — вектор параметров закона F_0.

Как правило, конкурирующую гипотезу формулируют так:

$$H_1 = \neg H_0$$

Для решения задачи проверки гипотезы H_0 против H_1 естественно сделать следующее:

1. Построить точечную оценку $\hat{\vec{\theta}}(\vec{X})$ для значения вектора параметров $\vec{\theta}$;

2. Использовать критерий Колмогорова для проверки простой гипотезы.

$$H_0 = \{F(t) \equiv F_0(t,\, \hat{\vec{\theta}}(\vec{x}))\}$$

где $\hat{\vec{\theta}}(\vec{x})$ — выборочное значение построенной оценки.

Недостатком данного подхода является то, что в этом случае критерий перестаёт быть параметрическим, т. к. распределение модифицированной статистики.

$$\triangle(\vec{X}) = \sup_{t \in \mathbb{R}} |F(t) - F_0(t,\, \hat{\vec{\theta}}(\vec{x}))|$$

зависит от выбранной точечной оценки, т. е. от закона распределения случайной величины $\hat{\vec{\theta}}(\vec{x})$. Однако можно показать, что если

1. $\hat{\vec{\theta}}(\vec{X})$ — оценка максимально правдоподобна для вектора $\vec{\theta}$,

2. Элементы $F_0(t,\, \vec{\theta})$ параметрического семейства получаются из какого-нибудь одного представителя с использованием преобразований сдвига и масштаба (вдоль оси Ot), т. е.
$$F_0(t,\, \vec{\theta}) = \tilde{F}_0(\tfrac{t-a}{b})$$
где $\tilde{F}_0(t)$ — какая-то фиксированная функция из рассматриваемого семейства $F_0(t,\, \vec{\theta})$; $a,\ b$ — компоненты, значения которых зависят от значения $\vec{\theta}$ в левой части,

то для исполнения критерия Колмогорова достаточно иметь только одну таблицу квантилей для каждого семейства (т. е. одну таблицу для нормального семейства, одну для экспоненциального и т. д.) Также таблицы составлены для наиболее важных семейств.

3.6.5 Критерий хи-квадрат для сложной гипотезы

Пусть

1. X — дискретная случайная величина;

2. X — может принимать значения из множества $a_1,\, \ldots,\, a_l$ с неизвестными вероятностями $p_1,\, \ldots,\, p_l$;

3. Эти вероятности p_k, $k = \overline{1;\, l}$ зависят от неизвестных параметров $\vec{\theta}$, где $\vec{\theta} \in \Theta$, т. е. в отличие от критерия χ^2 для простой гипотезы теперь $p_n = p_k(\vec{\theta})$, $\vec{\theta} \in \Theta$, $k = \overline{1;\, l}$.

По этой причине основную гипотезу можно записать в виде

$$H_0 = \{P\{X = a_k\} = p_{k_0}(\vec{\theta}),\, k = \overline{1;\, l}\}$$

Т. е. $p_{k_0}(\theta)$ — известные функции, предполагаемые зависимости вероятностей p_k от параметров $\vec{\theta}$; $P\{X = a_k\} = p_k(\vec{\theta})$ — теоретические ("истинные") зависимости этих вероятностей от параметров, эти зависимости нам неизвестны.

Конкурирующую гипотезу обычно выбирают как $H_1 = \neg H_0$; для решения задачи проверки H_0 против H_1 используется модифицированный критерий хи-квадрат.

1. Сначала строят оценку максимального правдоподобия для вектора параметров $\vec{\theta}$: $\hat{\vec{\theta}}(\vec{X})$;

2. Вычисляют выборочное значение $\hat{\vec{\theta}}(\vec{x})$, $n_k(\vec{x})$ — (см. критерий χ^2 для простой гипотезы);

3. Рассматривают статистику

$$\chi^2(\vec{X}) = \sum_{k=1}^{l} \frac{[n_k(\vec{X}) - np_{k_0}(\hat{\vec{\theta}}(\vec{X}))]^2}{np_{k_0}(\hat{\vec{\theta}}(\vec{X}))} = n \sum_{k=1}^{l} \frac{[\frac{n_k(\vec{X})}{n} - p_{k_0}(\hat{\vec{\theta}}(\vec{X}))]^2}{p_{k_0}(\vec{\theta}(\vec{X}))}$$

   которая, в случае выполнения определённых условий гладкости функции $p_{k_0}(\vec{\theta})$ при $n \to \infty$ слабо сходится к случайной величине, имеющей распределения $\chi^2(l-r-1)$, где r — размерность вектора $\vec{\theta}$.

4. Поскольку при истинности основной гипотезы H_0 статистика $\chi^2(\vec{X})$ принимает «малые» значения, а при истинности H_1 — «большие» значения, критическое множество можно записать в следующем виде:

$$W = \{\vec{x}\colon \chi^2(x^2) \geqslant h_{1-\alpha}^{(l-r-1)}\}$$

   где $h_{1-\alpha}^{(l-r-1)}$ — квантиль уровня $1-\alpha$ распределения $\chi^2(l-r-1)$.

Замечание. ***О построении оценки максимального правдоподобия в рассматриваемом случае.***

При истинности основной гипотезы H_0 функция правдоподобия имеет следующий вид:

$$L(\vec{X},\,\vec{\theta}) = \prod_{j=1}^{n} P\{X = X_j\} =$$

$$= \langle \text{т. к. среди компонент вектора } \vec{X} \text{ значение } a_k \text{ встречается } n_k(\vec{X}) \text{ раз, а элементы вектора } \vec{X} \text{ могут быть различными способами раскиданы по позициям в этом векторе, то это выражение можно записать в следующем виде} \rangle =$$

$$= \frac{n!}{n_1! \cdot \ldots \cdot n_k!} \prod_{k=1}^{l} \left[p_{k_0}(\vec{\theta})\right]^{h_k(\vec{X})}$$

где $\sum_{k=1}^{l} n_k(\vec{X}) = n$. Тогда уравнения правдоподобия

$$\frac{\delta \ln L}{\delta \theta_j} = 0,\ j = \overline{1;\, r}$$

примут вид

$$\sum_{k=1}^{l} \frac{n_k(\vec{X})}{p_{k_0}(\vec{\theta})} \cdot \frac{\delta p_{k_0}(\vec{\theta})}{\delta \theta_j} = 0,\ j = \overline{1;\, r}$$

Лекция №30, 13.05.2019

3.6.6 Двухвыборочная задача

Пусть

1. X, Y — случайные величины;
2. $F(t)$ — функция распределения случайной величины X, $G(t)$ — функция распределения случайной величины Y.
3. $\overline{X}$ — случайная выборка из генеральной совокупности X (объёма n_1), $\overline{Y}$ — случайная выборка из генеральной совокупности Y (объёма n_2).

Требуется проверить гипотезу

$$H_) = \{X \text{ и } Y \text{ одинаково распределены}\} = \{(\forall t \in \mathbb{R})(F(t) = G(t))\}$$

против $H_1 = \neg H_0$. Если случайные величины X и Y непрерывны, то для решения этой задачи можно использовать статистику $\triangle(\vec{X},\,\vec{Y})$, выборочные значения которой выражаются формулой

$$\triangle(\vec{x},\,\vec{y}) = \sup_{t \in \mathbb{R}} |F_{n_1}(t) - G_{n_2}(t)|$$

где $F_{n_1}(t)$, $G_{n_2}(t)$ — эмпирические функции распределения, отвечающие выборкам $\vec{x}$ и $\vec{y}$. Если истина H_0, то в соответствии с теоремой о сходимости функции распределения заключаем, что при достаточно больших n_1 и n_2 значения статистики должны быть «малыми», а при истинности H_1 — «большими». По этой причине критическое множество можно задать в виде:

$$W = \{(\vec{x}, \vec{y})\colon \triangle(\vec{x}, \vec{y}) \geqslant S_{1-\alpha}\}$$

где $\alpha \in (0;\, 1)$ — заданный уровень значимости критерия; $S_{1-\alpha}$ — квантиль уровня $1-\alpha$ закона распределения статистики $\triangle$ при истинности H_0.

Замечание 1. *Построенный критерий называется критерием Смирнова.*

Замечание 2. *О законе распределения статистки $\triangle(\vec{X}, \vec{Y})$.*

1. *Доказано, что* при истинности H_0 *закон распределения статистики $\triangle$ не зависит от $F(t)$ — теоретического (т. е. «истинного» закона) закона, распределения случайной величины X (поскольку H_0 предполагается истинной, X и Y имеет одинаковые распределения);*

2. *Для небольших n_1 и n_2 соответствующие распределения табулированы (т. е. составлены таблица их квантилей);*

3. *Смирнов доказал, что для $t > 0$*

$$P\{\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \triangle(\vec{X}, \vec{Y}) < t\} \xrightarrow[n_1\to\infty,\, n_2\to\infty]{} k(t)$$

где $k(t) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2t^2}$. Другими словами, при достаточно больших n_1 и n_2 можно считать, что случайная величина

$$A = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \triangle(\vec{X}, \vec{Y})$$

имеет своей функцией распределения $k(t)$, $t > 0$ (т. к. из определения статистики $\triangle$ вытекает, что $\triangle \geqslant 0$, то $A \geqslant 0 \implies F_A(t) \equiv 0$, $t \leqslant 0$).

Замечание 3. *О вычислении значений статистики $\triangle$.*

1. *Рассмотрим вариационный ряд $z_{(1)}, \ldots, z_{n_1+n_2}$ объединённой выборки*

$$(X_1, \ldots, X_{n_1}, Y_1, \ldots, \ldots, Y_{n_2})$$

Можно показать, что

$$\triangle(\vec{X}, \vec{Y}) = \max_{i=\overline{1;\, n}} |F_{n_1}(z_{(i)}) - G_{n_2}(z_{(i)})|$$

где $N = n_1 + n_2$.

2. *Обозначим*

$$d_i = \begin{cases} 1, & z_{(i)} \text{ — одно из наблюдений случайной величины} X \\ 0, & z_{(i)} \text{ — -=-} \end{cases}$$

(здесь $z_{(i)}$ — реализация $Z_{(i)}$).

Обозначим также

$$S_j = \frac{j \cdot n_1}{N} - \sum_{k=1}^{j} d_k, \ j = \overline{1;\, N}$$

Тогда

$$\triangle(\vec{x},\, \vec{y}) = \frac{N}{n_1 n_2} \max\{S_1,\, \ldots,\, S_n\}$$

Замечание. *В этом разделе мы изучили некоторые критерии согласия, которые используются для проверки гипотезы о том, что закон распределения случайной величины X имеет заданный вид.*

Другую большую группу критериев для проверки непараметрических гипотез составляет так называемые критерии независимости. Они предназначены для решения следующей задачи:

Дано:

1. *$(X,\, Y)$ — случайный вектор;*

2. *$(\vec{X},\, \vec{Y})$ — случайная выборка.*

Требуется проверить гипотезу $H_0 = \{X \text{ и } Y \text{ независимые}\}$ против $H_1 = \neg H_0$.

3.7 Элементы регрессионного анализа

3.7.1 Основные определения

Пример. *Пусть Y_1 — объём выпуска некоторым предприятием готовой продукции (млн. изделий в годы), а Y_2 — процент брака в этой продукции. Руководство предприятия считает, что величины Y_1 и Y_2 в основном зависят от следующих «параметров»:*

1. *X_1 — объём заработной платы (млн. руб.);*

2. *X_2 — объём вложений в социальный пакет (млн. руб.);*

3. *X_3 — объём вложений в подготовку/переподготовку (повышение квалификации кадров) (млн. руб.).*

Для оптимизации производственного процесса руководство предприятия хочет знать зависимость величин Y_1, Y_2 от величин X_1, X_2, X_3, т. е. хочет знать зависимости $Y_j = Y_j(X_1,\, X_2,\, X_3)$, $j = \overline{1;\, 2}$.

Особенностями данного примера являются следующие:

1. Величины X_1, X_2, X_3 являются детерминированными (т. е. не случайными);

2. Величины Y_1, Y_2 являются случайными.

Последнее объясняется тем, что на значения Y_1, Y_2 могут влиять неучтённые факторы (например, качество сырья, кол-во аварий, вложения в технику безопасности). Кроме этого, случайный характер величин Y_1 и Y_2 может быть связан с ошибками измерения (например, используется упрощённая схема контроля за качеством готовой продукции).

В этом случае связь между величинами Y_1, Y_2 и величинами X_1, X_2, X_3 является не функциональной, а стохастической (т. е. случайной).

Определение. *Говорят, что переменная Y стохастически зависит от переменных $X_1, \ldots, X_p$, если на изменения значений переменных $X_1, \ldots, X_p$ величина Y реагирует изменением своего закона распределения.*

Пример. *Пусть Y — случайная величина, функция плотности распределения которой имеет вид*

$$f_Y(y) = \frac{1}{\sqrt{2\pi}X_2} e^{-\frac{(y-X_1)^2}{2X_2^2}},\ y \in \mathbb{R}$$

В этом случае Y стохастически зависит от X_1, X_2.

Задачи, связанные с установкой стохастических зависимостей между случайной величиной Y и детерминизированными величинами $X_1, \ldots, X_p$, носящими количественный характер, составляет предмет изучения регрессионного анализа.

В регрессионном анализе используют модель чёрного ящика как наиболее общую модель отображения.

(картинка с чёрным ящиком, в которую идут стрелки от X_* и ε_* и из которого выходят стрелки в Y_*)

При этом используется терминология Ф-отображений, осуществляемых чёрным ящиком. Здесь $X_1, \ldots, X_p$ — факторы, $Y_1, \ldots, Y_s$ — отклики, $\varepsilon_1, \ldots, \varepsilon_m$ — случайные возмущения.

Вектор $\varepsilon_x = \vec{Y} - \Phi(\vec{X})$ называется вектором случайных ошибок.

3.7.2 Метод наименьших квадратов

Рассмотрим частный случай отображения Ф, в котором $p = m = s = 1$. Пусть у нас имеются результаты n наблюдений. $t_1, \ldots, t_n$ — n значений фактора X, а $y_1, \ldots, y_n$ — n отвечающих им откликов. Тогда

$$\begin{cases} y_1 = \Phi(t_1) + \varepsilon_1 \\ \ldots \\ y_n = \Phi(t_n) + \varepsilon_n \end{cases}$$

где $\varepsilon_1, \ldots, \varepsilon_n$ — реализация случайной ошибки ε.

Замечание 1. *Здесь ε — случайная величина, а $\varepsilon_1, \ldots, \varepsilon_n$ — её реализации.*

Замечание 2. *Рассматриваемая модель имеет вид $Y = \Phi(X) + \varepsilon$. В функции Φ нет ничего случайного, т. е. если известно значение X, то ему отвечает вполне определённое значение $\Phi(x)$.*

При этом сама функция Φ неизвестна неизвестна. Требуется по имеющимся результатам наблюдений подобрать некоторую функцию $\hat{\Phi}$ так, чтобы она наилучшим (в некотором смысле) образом аппроксимировала функцию $\Phi(x)$.

Лекция №31, 20.05.2019

1. Часто функцию $\hat{\Phi}$ ищут в виде

$$\hat{\Phi}(t) = \theta_1\psi_1(t) + \ldots + \theta_p\psi_p(t) \tag{30}$$

   где $\psi_j(t)$ — известные функции, которые часто называют базовыми; θ_j, $j = \overline{1; p}$ — неизвестные параметры.

   В этом случае задача тестирования построения отображения $\hat{\Phi}$ сводится к задаче подбора значений коэффициентов $\theta_1, \ldots, \theta_p$.

   Замечание 1. *Модель 30 называют линейной по параметрам, т. к. каждый параметр θ_j входит в правую часть линейно.*

   Замечание 2. *Задача выбора базисных функций $\psi_j(t)$, $j = \overline{1; n}$, является плохо формулированной и успех её рассмотрения часто зависит от интуиции и опыта исследователей.*

2. С учётом 30 результаты наблюдений можно записать в виде:

$$\begin{cases} y_1 = \theta_1\psi_1(t_1) + \ldots + \theta_p\psi_p(t_1) + \epsilon_1 \\ \ldots \\ y_n = \theta_1\psi_1(t_n) + \ldots + \theta_p\psi_p(t_n) + \epsilon_n \end{cases} \tag{31}$$

   Используем обозначения:

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \; \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \; \vec{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}, \; \Psi = \begin{bmatrix} \psi 1(t_1) & \ldots & \psi_p(t_1) \\ \vdots & \ddots & \vdots \\ \psi_1(t_n) & \ldots & \psi_p(t_n) \end{bmatrix}$$

   ([47])

   Чем более удачно подобран вектор $\vec{\theta}$, тем меньше будет значения функционала? $S(\vec{\theta})$.

[47] Ψ — прописная буква «пси» греческого алфавита. — Прим. ред.

Замечание.

$$S(\vec{\theta}) = \begin{bmatrix} y_1 - \hat{\Phi}(t_1) \\ \vdots y_n - \hat{\Phi}(t_n) \end{bmatrix}^\intercal \begin{bmatrix} y_1 - \hat{\Phi}(t_1) \\ \vdots y_n - \hat{\Phi}(t_n) \end{bmatrix} = (\vec{y} - \Psi\vec{\theta})^\intercal(\vec{y} - \Psi\vec{\theta}) = \|\vec{y} - \Psi\vec{\theta}\|^2$$

где $\|\cdot\|$ — евклидова норма.

Определение. *Оценкой, полученной по методу наименьших квадратов (МНК-оценкой) вектора $\vec{\theta}$ называют такое его значение $\hat{\vec{\theta}}$, которое доставляет наименьшее значение функционалу $S(\vec{\theta})$, т. е.*

$$S(\hat{\vec{\theta}}) = \min_{\vec{\theta}\in\mathbb{R}^p} S(\vec{\theta})$$

3. Как найти эту МНК-оценку?

Теорема. *Пусть $\operatorname{rg}\Psi = p$. ([48]) Тогда $\hat{\vec{\theta}} = (\psi^\intercal\psi)^{-1}\psi^\intercal\vec{y}$.*

Теорема. ***О свойствах МНК-оценки.***

Пусть

(a) $\varepsilon \sim N(0, \sigma^2)$, $i = \overline{1;\, n}$;

(b) *Реализация случайной величины ϵ в серии из n наблюдений независима.*

(c) $\operatorname{rg}\Psi = p$;

(d) $\hat{\theta} = (\psi^\intercal\psi)^{-1}\psi^\intercal\vec{y}$ — *МНК-оценка для $\vec{\theta}$.*

Тогда

(a) *$\hat{\theta}$ является несмещённой и состоятельной оценкой для $\vec{\theta}$ (из несмещённости следует $M[\hat{\vec{\theta}}] = \vec{\theta}$);*

(b) *$\vec{\theta} \sim N(\vec{\theta}, \Sigma)$ — нормальный случайный вектор, где $\Sigma = \sigma^2(\psi^\intercal\psi)^{-1}$;*

(c) *Интервальная оценка уровня $1-\alpha$ для параметра θ_j имеет вид $(\hat{\theta}_j - \triangle_j,\, \hat{\theta}_j + \triangle_j)$, где $\triangle_j = t_{1-\alpha}^{(n-p)}\sqrt{\frac{d_j}{n-p}S(\hat{\vec{\theta}})}$, $j = \overline{1;\, p}$; $t_{1-\alpha}^{(n-p)}$ — квантиль уровня $1-\alpha$ распределения $St(n-p)$, d_j — j-ый элемент главной диагонали матрицы $(\psi^\intercal\psi)^{-1}$.*

[48] $\operatorname{rg} X$ — ранг матрицы X. —Прим. ред.

A Комбинаторика

Пусть X — некое множество. Для примеров определим $X = \{1, 2, 3, 4, 5, 6\}$.

A.1 Сочетания без повторений

Определение. *Сочетанием без повторений из n ($n = |X|$) элементов по m называется любое неупорядоченное подмножество множества X, содержащее m различных элементов.*

Кол-во таких подмножеств обозначается как C_n^m и равно

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

A.2 Размещения без повторений

Определение. *Размещением без повторений из n элементов (исходного множества, $n = |X|$) по m (длина кортежа) называется кортеж, состоящий из m различных элементов множества X.*

Примеры. $(1, 2, 4)$, *но не* $(5, 5, 4)$.

Кол-во возможных размещений без повторений:

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

A.3 Перестановки без повторений

Перестановки без повторений — крайний случай размещений без повторений.

Определение. *Перестановкой без повторений называют называют кортеж, состоящий из $n = |X|$ различных элементов множества X.*

Кол-во возможных перестановок без повторений:

$$P_n = A_n^n = n!$$

Замечание. *Три предыдущих понятия связаны как*

$$A_n^m = P_m C_n^m$$

А.4 Размещения с повторениями

Определение. *Размещением с повторениями из n ($n = |X|$) по m элементов называется любой элемент из $X^m = X \times X \times \ldots \times X$.*

Примеры. $(1,\, 2,\, 3,\, 4,\, 5),\ (1,\, 4,\, 4,\, 4,\, 2)$.

Кол-во возможных размещений с повторениями:

$$\widetilde{A}_n^m = n^m$$

А.5 Перестановки с повторениями

Определение. *Перестановкой с повторениями называют кортеж длины n из элементов множества X, в котором каждый элемент $x_i \in X$ повторяется n_i, $i = \overline{1, k}$ раз ($\sum_i^k n_i = n$).*

Кол-во возможных перестановок с повторениями:

$$P_n(n_1,\, n_2,\, \ldots,\, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!},\ n = n_1 + n_2 + \ldots + n_k$$

B Заметки редактора

B.1 Рекомендуемая литература

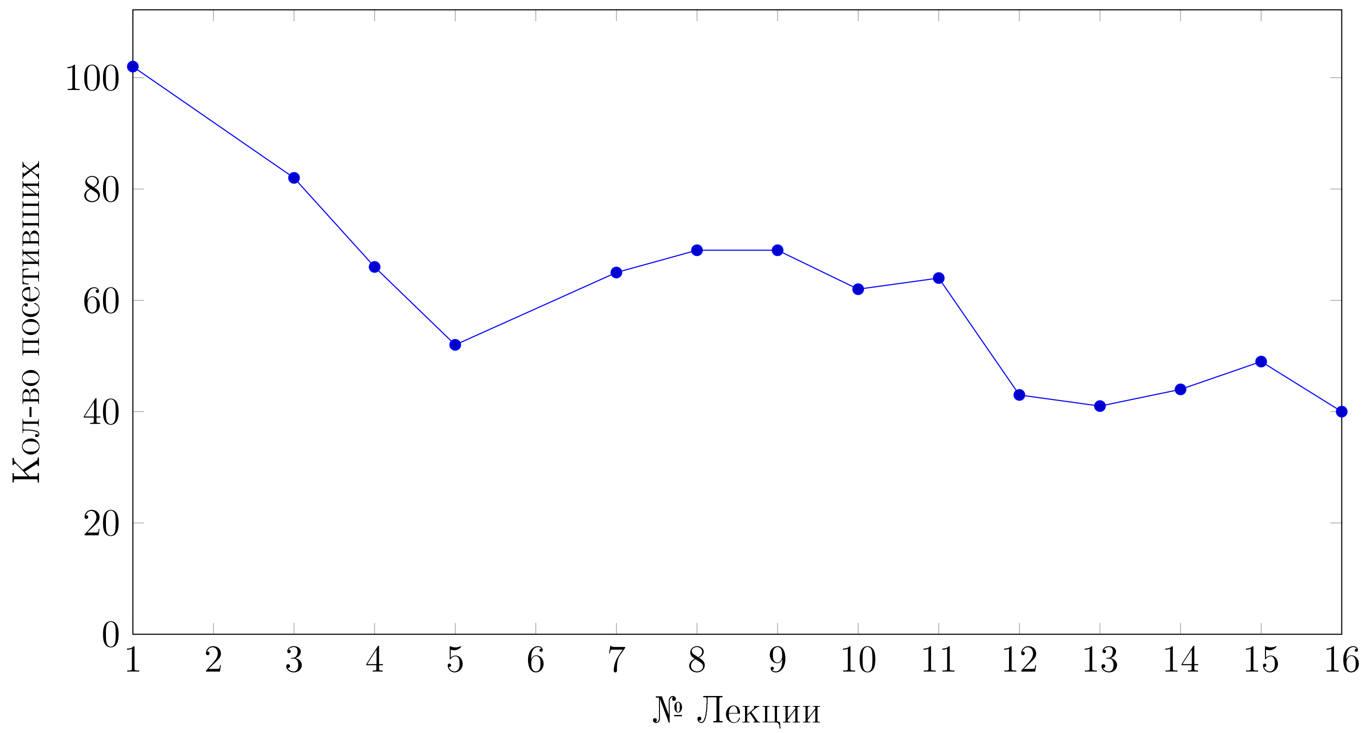
Для изучения Теории Вероятностей, параллельно с этими лекциями, можно читать:

- Очевидно, *«Теория Вероятностей»* из серии *«Математика в Техническом Университете»* (номер книги XVI). Данные лекции являются де-факто сокращённой версией этой книги;
- Саркисян П. С. (семинарист) рекомендует: *An Introduction to Probability Theory and its applications by William Feller*. Очень много хороших примеров... и опечаток;
- *A First Course in Probability by Sheldon M. Ross*, части 1-8. Курс примерно совпадает;
- Уровень «Хардкор», или книга для тех, кто хочет изучать *современную* теорию вероятностей: *Probability Theory: The Logic of Science by E. T. Jaynes.* Книга весьма слабо совпадает с описанным курсом: например, понятие вероятности в ней не строится на аксиомах Колмогорова, а выводится как численная функция степени правдоподобия утверждения.

И, да: восславь ~~солнце~~ Library Genesis (и ещё одна полезная ссылка).

C Статистические данные о средней посещаемости

C.1 Осенний семестр



C.2 Весенний семестр