Многомерные случайные величины (случайные векторы)

Двумерные случайные величины

Дискретные двумерные случайные величины.

Пусть X и Y — дискретные случайные величины, возможные значения которых $(x_i\ y_j)$, где $i=\overline{1,n}$; $j=\overline{1,m}$. Тогда распределение системы таких случайных величин может быть охарактеризовано указанием вероятностей

$$p_{ij}=P(X=x_i\,;\ Y=y_j)$$

того, что случайная величина X примет значение x_i и одновременно с этим случайная величина Y примет значение y_j.

Вероятности p_{ij} сводятся в таблицу вида

	x_1	x_2	…	x_n
y_1	p_{11}	p_{21}	…	p_{n1}
y_2	p_{12}	p_{22}	…	p_{n2}
⋮	⋮	⋮	…	⋮
y_m	p_{1m}	p_{2m}	…	p_{nm}

Все возможные события $(X=x_i;\ Y=y_j)$ при $i=\overline{1,n}$; $j=\overline{1,m}$ составляют полную группу несовместных событий, поэтому

$$\sum_{ij} p_{ij}=1=\sum_{ij} P(X=x_i;Y=y_j)$$

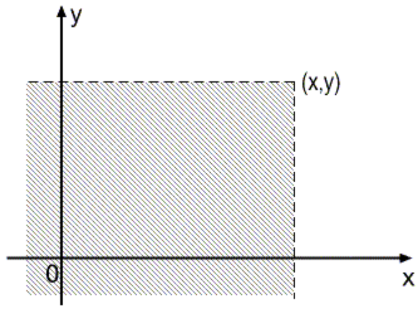
Зная закон распределения двумерной случайной величины, можно найти законы распределения ее составляющих (**частные законы**). Действительно, событие $X = x_1$ представляется собой сумму несовместных событий $(X = x_1,\ Y = y_1)$, $(X = x_1,\ Y = y_2)$,…, $(X = x_1,\ Y = y_m)$, поэтому $P(X = x_1) = P(x_1, y_1) + P(x_1, y_2) +\ldots+ P(x_1, y_m)$ (в правой части находится сумма вероятностей, стоящих в столбце, соответствующем $X = x_1$). Так же можно найти вероятности остальных возможных значений X. Для определения вероятностей возможных значений Y нужно сложить вероятности, стоящие в строке таблицы, соответствующей $Y = y_j$:

$$\sum_{j} p_{ij}=\sum_{j} P(X=x_i;Y=y_j)=P(X=x_i);$$

$$\sum_{i} p_{ij}=\sum_{i} P(X=x_i;Y=y_j)=P(Y=y_j).$$

Функция распределения и плотность распределения двумерной случайной величины

Определение: **Функцией распределения $F(x, y)$ двумерной случайной величины**



(X, Y) называется вероятность того, что $X < x$, а $Y < y$: $$F(x, y) = P(X < x, Y < y).$$

Геометрически функция распределения системы двух случайных величин представляет собой вероятность попадания случайной точки (X,Y) в левый нижний бесконечный квадрант плоскости с вершиной в точке (x, y).

Замечание: Определение функции распределения справедливо как для непрерывной, так и для дискретной двумерной случайной величины.

Свойства функции распределения:

1) $0 \le F(x, y) \le 1$ (так как $F(x, y)$ является вероятностью).

2) $F(x, y)$ есть неубывающая функция по каждому аргументу:

$F(x_2, y) \ge F(x_1, y)$, если $x_2 > x_1$; $F(x, y_2) \ge F(x, y_1)$, если $y_2 > y_1$.

Доказательство:

$F(x_2, y) = P(X < x_2, Y < y) = P(X < x_1, Y < y) + P(x_1 \le X < x_2, Y < y) \ge$

$\ge P(X < x_1, Y < y) = F(x_1, y)$. Аналогично доказывается и второе утверждение.

3) Имеют место предельные соотношения:

a) $F(-\infty, y) = 0$; b) $F(x, -\infty) = 0$; c) $F(-\infty, -\infty) = 0$; d) $F(\infty, \infty) = 1$.

Доказательство. События a), b) и c) невозможны (так как невозможно событие $X<-\infty$ или $Y<-\infty$), а событие d) достоверно, откуда следует справедливость приведенных равенств.

4) При $y = \infty$ функция распределения двумерной случайной величины становится функцией распределения составляющей X: $F(x, \infty) = F_1(x)$.

При $x = \infty$ функция распределения двумерной случайной величины становится функцией распределения составляющей Y: $F(\infty, y) = F_2(y)$.

Доказательство. Так как событие $Y < \infty$ достоверно, то $F(x, \infty) = P(X < x) = F_1(x)$. Аналогично доказывается второе утверждение.

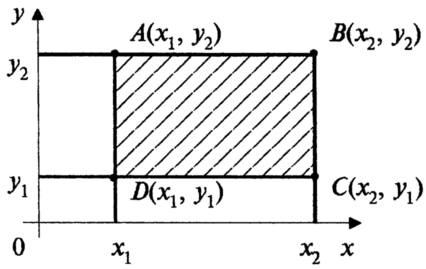
Или это свойство функции распределения *F(x, y)* можно проиллюстрировать, смещая ту или иную границу квадранта на $+\infty$. В этом случае квадрант превращается в полуплоскость, вероятность попадания в которую есть функция распределения случайной величины, соответствующей тому аргументу, который **не** равен $+\infty$.

Вероятность попадания в полу полосу $P(x_1 \le X < x_2; Y < y) = F(x_2; y) - F(x_1; y)$,

аналогично $P(X < x; y_1 \le Y < y_2) = F(x_1; y_2) - F(x_1; y_1)$.

Вероятность попадания в прямоугольник со сторонами, параллельными координатным осям, равна

$P(x_1 \le X < x_2; y_1 \le Y < y_2) = [F(x_2; y_2) - F(x_1; y_2] - [F(x_2; y_1) - F(x_1; y_1)]$.



Система двух случайных величин *(X, Y)* называется **непрерывной**, если ее функция распределения *F(x, y)* есть непрерывная функция, дифференцируемая по каждому из аргументов, у которой существует вторая смешанная производная $\frac{\partial^2 F(x,y)}{\partial x \partial y}$. Обе составляющие системы X и Y представляют собой непрерывные случайные величины.

Определим плотность распределения системы двух величин аналогично тому, как мы определили плотность распределения для одной случайной величины.

Пусть имеется система двух непрерывных случайных величин *(X, Y)*. Рассмотрим вероятность попадания случайной величины *(X, Y)* в элементарный прямоугольник со сторонами Δx и Δy, примыкающий к точке с координатами *(x, y)*

Тогда имеем

$$P(x<X<x+\Delta x;\ y<Y<y+\Delta y)=F(x+\Delta x;\ y+\Delta y) - F(x;\ y+\Delta y) - (F(x+\Delta x;\ y)-F(x;\ y)).$$

Разделим полученную вероятность на площадь этого прямоугольника и перейдем к пределу при $\Delta x \to 0$ и $\Delta y \to 0$:

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P(x < X < x + \Delta x; y < Y < y + \Delta y)}{\Delta x \Delta y} =$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F(x + \Delta x; y + \Delta y) - F(x; y + \Delta y) - F(x + \Delta x; y) + F(x; y)}{\Delta x \cdot \Delta y}$$

Предположим, что функция распределения *F(x, y)* не только непрерывна, но и дважды дифференцируема. Тогда правая часть последнего равенства представляет собой вторую смешанную частную производную функции *F(x, y)*. Обозначим эту производную через *f(x, y)*:

$$f(x, y)= \frac{\partial^2 F(x,y)}{\partial x \partial y} = F''_{xy}\ (x, y).$$

Функция *f(x, y)* называется **плотностью распределения** системы непрерывных случайных величин *(X, Y)*.

Плотность *f(x, y)* обладает следующими свойствами:

1) $f(x, y) \geq 0$ следует из того, что $F(x, y)$ — неубывающая функция по каждому из аргументов

2) $F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dxdy$ - следует из определения двумерной плотности вероятности.

3) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$ поскольку это вероятность того, что точка попадет на плоскость Оxy, то есть достоверного события.

Вероятность попадания случайной точки в произвольную область

Пусть в плоскости Оху задана произвольная область D. Найдем вероятность того, что точка, координаты которой представляют собой систему двух случайных величин (двумерную случайную величину) с плотностью распределения $f(x, y)$, попадет в область D. Разобьем эту область прямыми, параллельными осям координат, на прямоугольники со сторонами Δx и Δy. Вероятность попадания в каждый такой прямоугольник равна $f(\xi_i, \eta_i)\Delta x \Delta y$, где (ξ_i, η_i) - координаты точки, принадлежащей прямоугольнику. Тогда вероятность попадания точки в область D есть предел интегральной суммы $\sum_{i=1}^{n} f(\xi_i, \eta_i)\Delta x \Delta y$, то есть

$$P((X, Y) \in D) = \iint_D f(x,y)\,dx\,dy.$$

По известной плотности совместного распределения можно найти плотности распределения каждой из составляющих двумерной случайной величины.

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy; \qquad f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Условные законы распределения

Определение: Распределение одной случайной величины, входящей в систему, найденное при условии, что другая случайная величина приняла определенное значение, называется **условным законом распределения**.

Рассмотрим дискретную двумерную случайную величину и найдем закон распределения составляющей X при условии, что Y примет определенное значение (например, $Y = y_1$). Для этого воспользуемся формулой Байеса, считая гипотезами события $X = x_1$, $X = x_2$,…, $X = x_n$, а событием A – событие $Y = y_1$. При такой постановке задачи нам требуется найти условные вероятности гипотез при условии, что A произошло. Следовательно,

$$p(x_i / y_1) = \frac{p(x_i, y_1)}{p(y_1)}.$$

Таким же образом можно найти вероятности возможных значений X при условии, что Y принимает любое другое свое возможное значение:

$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)}.$$

Аналогично находят условные законы распределения составляющей *Y*:

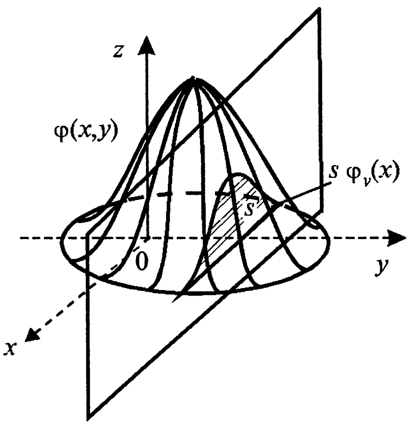
$$p(y_j / x_i) = \frac{p(x_i, y_j)}{p(x_i)}.$$

Условный закон распределения можно задавать как функцией распределения так и плотностью распределения.

Условная плотность распределения вычисляется по формулам:

$$f(x/y) = \frac{f(x,y)}{f_2(y)} = \frac{f(x,y)}{\int\limits_{-\infty}^{\infty} f(x,y)dx} \qquad f(y/x) = \frac{f(x,y)}{f_1(x)} = \frac{f(x,y)}{\int\limits_{-\infty}^{\infty} f(x,y)dy}$$

Условная плотность распределения обладает всеми свойствами плотности распределения одной случайной величины.



Если плотность *f(x, y)* представляет собой геометрически некоторую поверхность распределения, то условная плотность *f(x/y)* есть кривая распределения, **подобная** сечению этой поверхности плоскостью *Y=y* , параллельной координатной плоскости *OXZ* (на рисунке $\varphi(x, y)$ и $\varphi_y(x)$ - плотность и условная плотность).

Пример: Случайная точка *(X, Y)* попадает с одинаковой вероятностью в любую точку эллипса с полуосями *a, b*, совпадающими с осями координат. Определить плотность вероятности случайных величин *X* и *Y*, а также условные плотности вероятности *f(y / x)* и *f(x / y)*.

Решение: Если двумерная случайная величина равномерно распределена в области *G*, площадь которой равна *Q* , то плотность вероятности находится следующим образом.

Из условия примера ясно, что плотность вероятности следует считать постоянной. Пусть *f(x, y)=A*. Тогда согласно нормировочному свойству

$$\iint\limits_G A\,dx\,dy = 1$$

По известному свойству двойных интегралов

$$\iint\limits_G dx\,dy = Q \qquad \Rightarrow \qquad A = \frac{1}{Q}. \qquad f(x, y) = \frac{1}{Q}$$

для $(X, Y) \in G$, где *Q* — площадь области *G*. Но площадь эллипса равна $Q = \pi ab$. Следовательно, в нашем случае, имеем

$$f(x,y) = \begin{cases} \dfrac{1}{\pi ab}, & \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \le 1, \\ 0, & \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} > 1. \end{cases}$$

Если $x \in [-a; a]$, то

$$-b\sqrt{1-\frac{x^2}{a^2}} \le y \le b\sqrt{1-\frac{x^2}{a^2}}$$

Тогда, при $|x| \le a$

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)dy = \int_{-\frac{b}{a}\sqrt{a^2-x^2}}^{\frac{b}{a}\sqrt{a^2-x^2}} \frac{1}{\pi ab}dy = \frac{2}{\pi a^2}\sqrt{a^2-x^2}\,.$$

При $|x| > a$, $f_1(x)\equiv 0$.

Аналогично, при $|y| \le b$

$$f_2(y)=\int_{-\infty}^{\infty} f(x,y)dx = \int_{-\frac{a}{b}\sqrt{b^2-y^2}}^{\frac{a}{b}\sqrt{b^2-y^2}} \frac{1}{\pi ab}dx = \frac{2}{\pi b^2}\sqrt{b^2-y^2}\,.$$

При $|y| > b$, $f_2(y)\equiv 0$.

Найдем, далее $f(y/x)$:

$$f(y/x)=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy} = \frac{\frac{1}{\pi ab}}{\frac{2}{\pi a^2}\sqrt{a^2-x^2}} = \frac{a}{2b\sqrt{a^2-x^2}}$$

при $|x| < a$, и $|y| \le \frac{b}{a}\sqrt{a^2-x^2}$. При нарушении хотя бы одного из последних неравенств

$$f(y/x)\equiv 0.$$

Аналогично,

$$f(x/y)=\begin{cases} \dfrac{b}{2a\sqrt{b^2-y^2}}, & |y|<b;\quad |x| \le \dfrac{a}{b}\sqrt{b^2-y^2}, \\ 0, & \text{в противном случае.} \end{cases}$$

Определение: **Условным математическим ожиданием** дискретной случайной величины Y при X = x (x – определенное возможное значение X) называется произведение всех возможных значений Y на их условные вероятности

$$M(Y/X=x)=\sum_{i=1}^{m} y_i p(y_i/x)$$

Для непрерывных случайных величин: $M(Y/X=x)=\int_{-\infty}^{\infty} yf(y/x)dy$,

Где $f(y/x)$ – условная плотность случайной величины Y при $X=x$.

Условное математическое ожидание $M(Y/x)=f(x)$ является функцией от x и называется **функцией регрессии *Y* на *X*.**

Аналогично, можно записать $M(X/Y=y)=f(y)$ - т.е. это функция регрессии X на Y.

График той или иной функции называется **линией регрессии** или кривой регрессии.

Аналогично определяются условная дисперсия и условные моменты системы случайных величин. Например

$$D(Y/X=x)=\int_{-\infty}^{\infty}\left(y-M(Y/X=x)\right)^2 f(y/x)dy.$$

Зависимые и независимые случайные величины

Две случайные величины X и Y называются **независимыми**, если независимы все связанные с ними события: например,

$(X<x)$ и $(Y<y)$; $(X=x_i)$ и $(Y=y_i)$ и т. д.

В терминах законов распределения независимость случайных величин можно определить так: две случайных величины называются независимыми, если закон распределения каждой из них не зависит от того, какое значение приняла другая.

Если случайные величины X и Y, образующие систему, независимы, то функция распределения системы выражается через функции распределения отдельных величин, входящих в систему. В самом деле,

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

Но события $(X<x)$ и $(Y<y)$ независимы; по правилу умножения вероятностей для независимых событий

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

или

$$F(x, y)=F_1(x)\cdot F_2(y).$$

Определим необходимые и достаточные условия независимости случайных величин.

Теорема: Для того, чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функций распределения составляющих.

$$F(x,y)=F_1(x)F_2(y)$$

Аналогичную теорему можно сформулировать и для плотности распределения:

Теорема: Для того, чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы плотность совместного распределения системы (X, Y) была равна произведению плотностей распределения составляющих

$$f(x,y)=f_1(x)f_2(y).$$

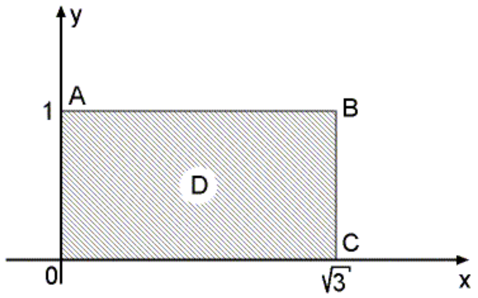
Используя этот результат, можно доказать свойство мат.ожидания (аналогично и для дискретного случая): математическое ожидание произведения двух **независимых** случайных величин равно произведению их математических ожиданий $M(XY)=M(X)M(Y)$.

$$M(XY)=\iint\limits_{-\infty}^{\infty} xyf(x,y)dxdy=\iint\limits_{-\infty}^{\infty} xyf_1(x)\cdot f_2(y)dxdy=\int\limits_{-\infty}^{\infty} xf_1(x)dx\int\limits_{-\infty}^{\infty} yf_2(y)dy=MX\cdot MY .$$

Как уже говорилось в предыдущем параграфе, если случайные величины, образующие систему, зависимы, то для нахождения закона распределения системы недостаточно знать законы распределения отдельных величин, входящих в систему: требуется еще знать условный закон распределения одной из них.

Пример: $f(x, y)=\dfrac{a}{1+x^2+x^2y^2+y^2}$

1. Определить, зависимы или нет случайные величины *X*, *Y*.
2. Найти *a*.
3. Найти *F(x, y)*.
4. Найти вероятность попадания случайных точек *(X, Y)* в прямоугольник *D*.



Решение:

1. $f(x,y)=\dfrac{a}{1+x^2+x^2y^2+y^2}=$

$$a\frac{1}{\left(1+x^2\right)\left(1+y^2\right)}= =a\frac{1}{1+x^2}\cdot\frac{1}{1+y^2}=af_1(x)\cdot f_2(y)$$

то есть случайные величины X и Y независимы.

2. Для определения *a* воспользуемся нормировочным свойством:

$$\iint\limits_{-\infty}^{\infty} f(x,y)dx\,dy \; = a\iint\limits_{-\infty}^{\infty}\frac{dx\,dy}{\left(1+x^2\right)\left(1+y^2\right)}=a\int\limits_{-\infty}^{\infty}\frac{dx}{1+x^2}\int\limits_{-\infty}^{\infty}\frac{dy}{1+y^2}=$$

$$=a\cdot\operatorname{arctg}x\Big|_{-\infty}^{\infty}\cdot\operatorname{arctg}y\Big|_{-\infty}^{\infty}=a\pi^2=1.$$

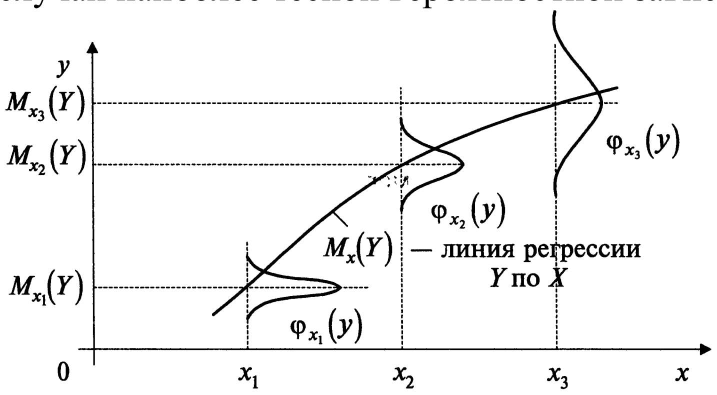
Следовательно, $a=\dfrac{1}{\pi^2}$.

3. $F(x, y)=\dfrac{1}{\pi^2}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y}\dfrac{dx\,dy}{\left(1+x^2\right)\left(1+y^2\right)}=\left(\dfrac{1}{\pi}\operatorname{arctg}x+\dfrac{1}{2}\right)\left(\dfrac{1}{\pi}\operatorname{arctg}y+\dfrac{1}{2}\right).$

4. $P((X, Y)\subset D)=\dfrac{1}{\pi^2}\iint\limits_{D}\dfrac{dx\,dy}{\left(1+x^2\right)\left(1+y^2\right)}=\dfrac{1}{\pi^2}\int\limits_{0}^{\sqrt{3}}\dfrac{dx}{1+x^2}\int\limits_{0}^{1}\dfrac{dy}{1+y^2}=$

$$=\frac{1}{\pi^2}\operatorname{arctg}x\Big|_{0}^{\sqrt{3}}\cdot\operatorname{arctg}y\Big|_{0}^{1}=\frac{1}{\pi^2}\cdot\frac{\pi}{3}\cdot\frac{\pi}{4}=\frac{1}{12}$$

До сих пор мы сталкивались с понятием функциональной зависимости между переменными X и Y, когда каждому значению x соответствовало строго определённое значение y. Зависимость между двумя случайными величинами называется **вероятностной** (**стохастической** или **статистической**), если каждому значению одной из них соответствует определённое (условное) распределение другой (т.е. нельзя, зная значение одной из них, точно определить значение другой, а можно указать лишь распределение её распределение). Например, рост и вес , рост и возраст человека. Функциональную зависимость можно рассматривать как крайний, предельный случай наиболее тесной вероятностной зависимости.



На рисунке ($M_x(Y)$ и $\varphi_x(y)$ условные мат.ожидания и условные плотности) зависимость между X и Y проявляется в том, что с изменением x меняется как распределение Y, так и условное мат.ожидание.

Коэффициент корреляции. Уравнение линейной регрессии

Определение : **Корреляционным моментом (ковариацией)** системы двух случайных величин называется:

$$K_{xy} = M\{(X - M(X))(Y - M(Y))\} = cov(X;Y)$$

$$K_{xy} = \sum_{i,j=1}^{n,m}(x_i - M(X))(y_j - M(Y))p_{ij} \ ,$$

$$K_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - M(X))(y - M(Y))f(x;y)dxdy \ .$$

Если раскрыть скобки, получим

$$M\{(X - M(X))(Y - M(Y))\} = M\{XY - XM(Y) - YM(X) + M(X)M(Y)\} =$$
$$= M(XY) - M(X)M(Y) - M(Y)M(X) + M(X)M(Y) = M(XY) - M(X)M(Y).$$

Из свойств MX следует также, что $cov\,(X, Y)= cov\,(Y{\cdot}X)$ и $cov\,(X;X)=D(X)$. Также отметим, что $D(X+Y)=DX+DY+2\,cov\,(X,Y)$.

Корреляционный момент служит для того, чтобы охарактеризовать связь между случайными величинами. Если случайные величины независимы, то их корреляционный момент равен нулю.

Корреляционный момент имеет размерность, равную произведению размерностей случайных величин X и Y. Этот факт является недостатком этой числовой характеристики, т.к. при различных единицах измерения получаются различные корреляционные моменты, что затрудняет сравнение корреляционных моментов различных случайных величин. Для того, чтобы устранить этот недостаток применятся другая характеристика – коэффициент корреляции.

Определение: **Коэффициентом корреляции** r_{xy} случайных величин X и Y называется отношение корреляционного момента к произведению средних квадратических отклонений этих величин.

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}$$

Коэффициент корреляции является безразмерной величиной. Коэффициент корреляции характеризует степень зависимости случайных величин X и Y, причём не любую, а только **линейную**. При этом положительная корреляция между случайными величинами означает, что при возрастании одной из них другая имеет тенденцию к возрастанию; отрицательная же корреляция означает, что при возрастании одной из случайных величин другая имеет тенденцию к убыванию.

Свойства:

1) Коэффициент корреляции независимых случайных величин равен нулю.
2) Абсолютная величина коэффициента корреляции не превышает единицы

$$|r_{xy}| \le 1.$$

Доказательство: Введём нормированные случайные величины:
$Z_1 = \dfrac{X - M(X)}{\sigma_x}$, $Z_2 = \dfrac{Y - M(Y)}{\sigma_y}$.

Тогда в соответствии с ранее написанным $D(Z_1 \pm Z_2) = D(Z_1) + D(Z_2) \pm 2cov(Z_1, Z_2)$.

Так как а) $D(Z_1 \pm Z_2) \ge 0$; б) $cov(Z_1, Z_2) \overset{\Delta}{=} M((Z_1 - MZ_1)(Z_2 - MZ_2)) =$

$$= M\left\{\left(\frac{X - MX}{\sigma_x} - M\left(\frac{X - MX}{\sigma_x}\right)\right)\left(\frac{Y - MY}{\sigma_y} - M\left(\frac{Y - MY}{\sigma_y}\right)\right)\right\} =$$

$$= \frac{1}{\sigma_x \sigma_y} M\{(X - MX - MX + MX)(Y - MY - MY + MY)\} =$$

$$= \frac{M\{(X - MX)(Y - MY)\}}{\sigma_x \sigma_y} = \frac{\mathrm{cov}(X, Y)}{\sigma_x \sigma_y} = r_{xy}.$$

в) $DZ_1 \overset{\Delta}{=} M(X - MX)^2 = M\left(\dfrac{X - MX}{\sigma_x} - M\left(\dfrac{X - MX}{\sigma_x}\right)\right)^2 = \dfrac{M(X - MX - MX + MX)}{\sigma_x^2} = \dfrac{DX}{\sigma_x^2} = 1$.

Аналогично $DZ_2 = 1$. Тогда $D(Z_1 \pm Z_2) = 2 \pm 2r_{xy} = 2(1 \pm r_{xy}) \ge 0$

$\Rightarrow \left.\begin{array}{l} 1 + r_{xy} \ge 0 \\ 1 - r_{xy} \ge 0 \end{array}\right\} \Rightarrow |r_{xy}| \le 1,$ что и требовалось доказать.

3) Если коэффициент корреляции двух случайных величин равен (по абсолютной величине) единице, то между этими случайными величинами существует линейная функциональная зависимость.

Доказательство: так как $D(Z_1 \pm Z_2) = 2 \pm 2r_{xy}$ при $r_{xy} = \pm 1$

$D(Z_1 \pm Z_2) = D\left(\dfrac{X - M(X)}{\sigma_x} \pm \dfrac{Y - M(Y)}{\sigma_y}\right) = 0$. Это означает, что $\dfrac{X - M(X)}{\sigma_x} \pm \dfrac{Y - M(Y)}{\sigma_y} = 0$.

При $r_{xy} = 1$ получим $Y = M(Y) + \dfrac{\sigma_y}{\sigma_x}(X - M(X))$,

при $r_{xy} = -1$ $Y = M(Y) - \dfrac{\sigma_y}{\sigma_x}(X - M(X))$, т.е. X и Y связаны линейной функциональной зависимостью.

Случайные величины называются **коррелированными**, если их корреляционный момент отличен от нуля, и **некоррелированными**, если их корреляционный момент равен нулю.

Если случайные величины независимы, то они и некоррелированы, но из некоррелированности нельзя сделать вывод о их независимости (отсутствует только линейная зависимость между случайными величинами; любой другой вид связи при этом может присутствовать) .

Если две величины зависимы, то они могут быть как коррелированными, так и некоррелированными.

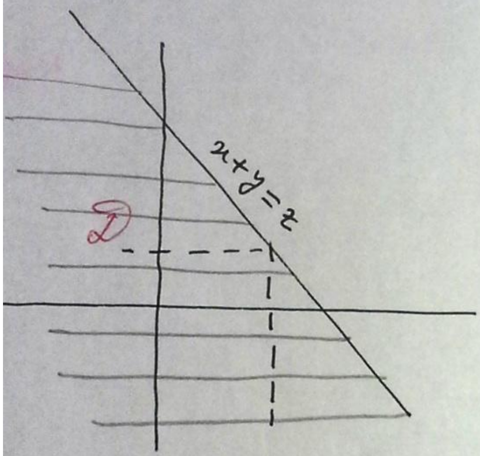
Задача композиции

$X: \ f_X(x)$, $\quad Y: \ f_Y(y)$

X и Y – независимые случайные величины

Найти функцию плотности распределения случайной величины

$$Z = X + Y$$



$$F(z) = P\{Z < z\} = P\{X + Y < z\}$$

$$f_Z(x, y) = f_X(x) f_Y(y)$$

$$F(z) = \iint\limits_{D} f_X(x) f_Y(y)\,dx\,dy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} f_X(x) f_Y(y)\,dy$$

$$f(z) = F'(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,dx$$

Пример (распределение хи-квадрат с двумя степенями свободы)

Пусть X и Y имеют стандартное нормальное распределение $X \propto N(0,1)$ и $Y \propto N(0,1)$.

$$X^2: \quad f_{X^2}(x) = \begin{cases} \dfrac{1}{2\sqrt{2\pi x}} e^{-x/2}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

$$Y^2: \quad f_{Y^2}(y) = \begin{cases} \dfrac{1}{2\sqrt{2\pi y}} e^{-y/2}, & y \ge 0 \\ 0, & y < 0 \end{cases}$$

$$Z = X^2 + Y^2$$

При $z > 0$

$$f_Z(z) = \int_0^z \frac{1}{2\sqrt{2\pi x}} e^{-x/2} \frac{1}{2\sqrt{2\pi x}} e^{-(z-x)/2} dx =$$

$$= \frac{1}{2\pi} e^{-z/2} \int_0^z \frac{dx}{\sqrt{xz - x^2}} = \frac{1}{2\pi} e^{-z/2} \int_0^z \frac{dx}{\sqrt{\frac{z^2}{2} - \left(x - \frac{z}{2}\right)^2}} =$$

$$= \frac{1}{2\pi} e^{-z/2} \arcsin \left. \frac{x - \frac{z}{2}}{z/2} \right|_0^z = \frac{1}{2\pi} e^{-z/2} \left(\arcsin(1) - \arcsin(-1) \right) = \frac{1}{2} e^{-z/2}$$

Распределение хи-квадрат

$$\chi^2(n) = U_1^2 + U_2^2 + .. + U_n^2$$

$$U_k \propto N(0,1)$$

$$f_U = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$f_{\chi^2(1)}(x) = \begin{cases} \frac{1}{2\sqrt{2\pi x}} e^{-x/2} & , x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$f_{\chi^2(n)}(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n-2)/2} e^{-x/2} & , x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$f_{\chi^2(n+1)}(z) = \int_0^z f_{\chi^2(n)}(x) f_{\chi^2(1)}(z-x)\, dx =$$

$$= \int_0^z \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n-2)/2} e^{-x/2} \frac{1}{2\sqrt{2\pi x}} e^{-(z-x)/2} dx =$$

$$= e^{-z/2} \frac{1}{2^{n/2}\Gamma(n/2)} \frac{1}{2\sqrt{2\pi}} \int_0^z x^{(n-2)/2} x^{-1/2} dx =$$

$$= e^{-z/2} \frac{1}{2^{n/2}\Gamma(n/2)} \frac{1}{2\sqrt{2\pi}} z^{(n-1)/2} \frac{1}{(n-1)/2}.$$

Обозначим числовой коэффициент

$$C=\frac{1}{2^{n/2}\Gamma(n/2)}\frac{1}{(n-1)/2}\cdot\frac{1}{2\sqrt{2\pi}}$$

Тогда для функции плотности получим выражение

$$f_{\chi^2(n+1)}(z)=C\, z^{(n-1)/2}\,\mathrm{e}^{-z/2}$$

Коэффициент C найдем из условия нормировки:

$$\int_0^\infty f_{\chi^2(n+1)}(z)\,dz=1$$

$$\int_0^\infty C\, z^{(n-1)/2}\,\mathrm{e}^{-z/2}\,dz=1$$

$$C\int_0^\infty z^{\frac{n+1}{2}-1}\,\mathrm{e}^{-z/2}\,dz=1$$

После замены переменной $t=\frac{z}{2}$ получим

$$C\cdot 2^{(n+1)/2}\int_0^\infty t^{\frac{n+1}{2}-1}\,\mathrm{e}^{-t}\,dt=1$$

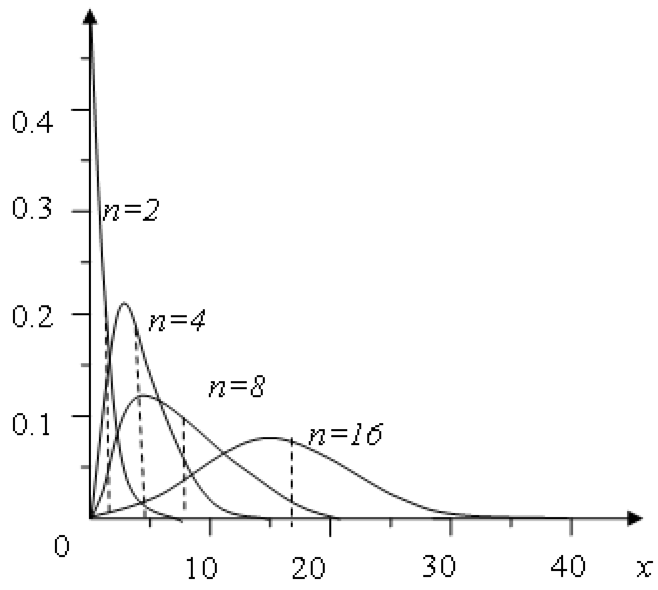
$$\Gamma(z)=\int_0^\infty x^{z-1}\,\mathrm{e}^{-x}\,dx$$

$$C\cdot 2^{(n+1)/2}\,\Gamma\left(\frac{n+1}{2}\right)=1$$

Отсюда имеем

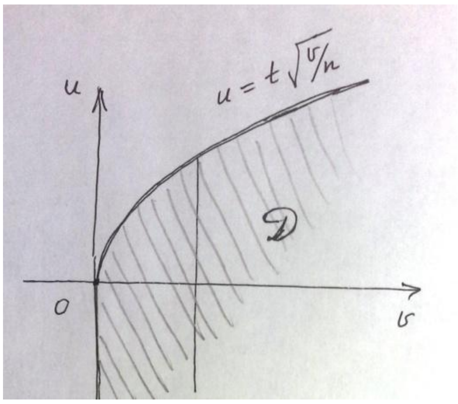
$$C=\frac{1}{2^{(n+1)/2}\,\Gamma\left(\frac{n+1}{2}\right)}$$

$$f_{\chi^2(n+1)}(x)=\frac{1}{2^{\frac{n+1}{2}}\,\Gamma\left(\frac{n+1}{2}\right)}\,x^{(n-1)/2}\,\mathrm{e}^{-x/2}\quad ,x\ge 0$$



$$f_{\chi^2(n)}(x) = \begin{cases} 0, \; x < 0 \\ \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{(n-2)/2}\, \mathrm{e}^{-x/2}, \; x \ge 0 \end{cases}$$

$$\Gamma(z) = \int_0^{\infty} x^{z-1}\, \mathrm{e}^{-x}\, dx$$



Распределение Стьюдента

$$T(n) = \frac{U}{\sqrt{\chi^2(n)/n}}$$

$$U = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

$$F(t) = P\{T < t\} = P\left\{ \frac{U}{\sqrt{V/n}} < t \right\} = \int_0^{\infty} dv \int_{-\infty}^{t\sqrt{v/n}} du \ f_U(u) f_V(v) =$$

$$= \int_0^{\infty} dv \int_{-\infty}^{t\sqrt{v/n}} du \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-u^2/2} \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n-2)/2} \mathrm{e}^{-v/2} =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} dv \int_{-\infty}^{t\sqrt{v/n}} du \, \mathrm{e}^{-u^2/2} \, v^{(n-2)/2} \mathrm{e}^{-v/2}$$

$$f(t) = F'(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} dv \, v^{(n-2)/2} \mathrm{e}^{-v/2} \frac{d}{dt} \int_{-\infty}^{t\sqrt{v/n}} du \, \mathrm{e}^{-u^2/2} =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} dv \, v^{(n-2)/2} \mathrm{e}^{-v/2} \sqrt{\frac{v}{n}} \, \mathrm{e}^{-\frac{v}{n}t^2} =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{n/2}\Gamma(n/2)} \sqrt{\frac{1}{n}} \int_0^{\infty} dv \, v^{(n-1)/2} \mathrm{e}^{-v/2} \mathrm{e}^{-\frac{v}{2}(1+t^2/n)}$$

Выполним замену переменной: $z=\frac{v}{2}\left(1+\frac{t^2}{n}\right)$.

$$f(t)=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty dv\, v^{(n-2)/2}\,e^{-v/2}\,\frac{d}{dt}\int_{-\infty}^{t\sqrt{v/n}}du\,e^{-u^2/2}=$$

$$=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty dv\, v^{(n-2)/2}\,e^{-v/2}\sqrt{\frac{v}{n}}\,e^{-\frac{v}{n}t^2}=$$

$$=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{n/2}\Gamma(n/2)}\sqrt{\frac{1}{n}}\int_0^\infty dv\, v^{(n-1)/2}\,e^{-v/2}e^{-\frac{v}{2}(1+t^2/n)}=$$

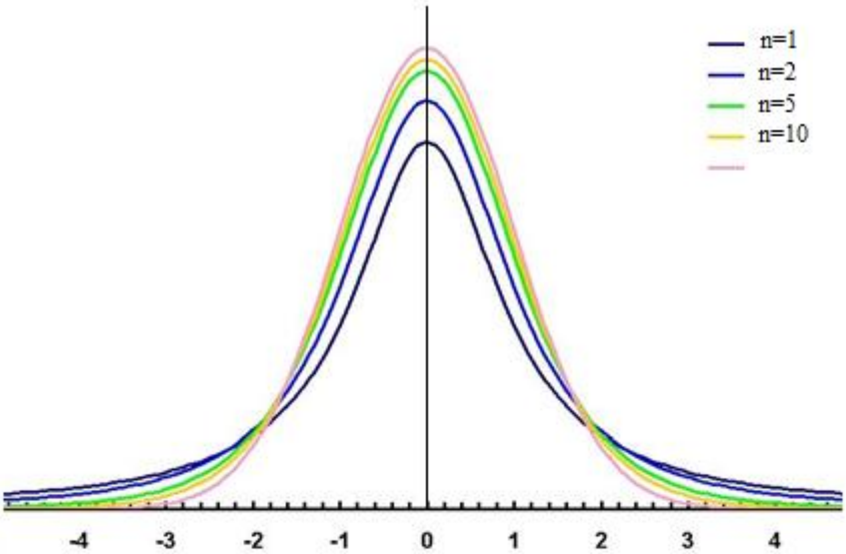
$$=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{n/2}\Gamma(n/2)}\frac{2^{(n+1)/2}}{\left(1+t^2/n\right)^{(n+1)/2}}\sqrt{\frac{1}{n}}\int_0^\infty dz\, z^{(n-1)/2}\,e^{-z}=$$

$$=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{n/2}\Gamma(n/2)}\frac{2^{(n+1)/2}}{\left(1+t^2/n\right)^{(n+1)/2}}\sqrt{\frac{1}{n}}\,\Gamma\left(\frac{n+1}{2}\right)=$$

$$=\frac{1}{\sqrt{n\pi}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)}\left(1+t^2/n\right)^{-(n+1)/2}$$

Окончательно:

$$f_{T(n)}(x)=\frac{1}{\sqrt{n\pi}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)}\left(1+x^2/n\right)^{-(n+1)/2}.$$

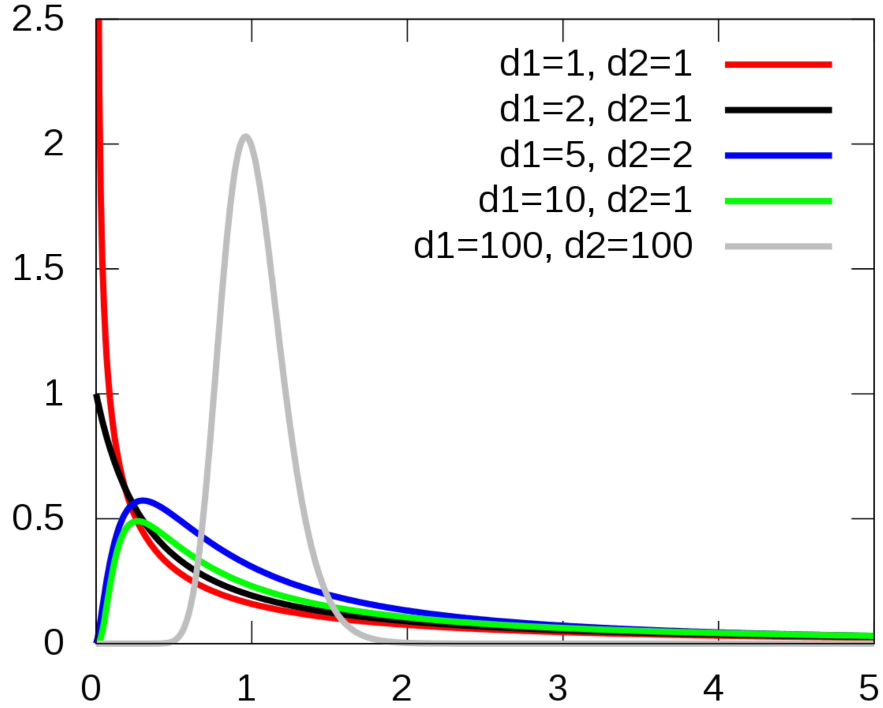


Распределение Фишера

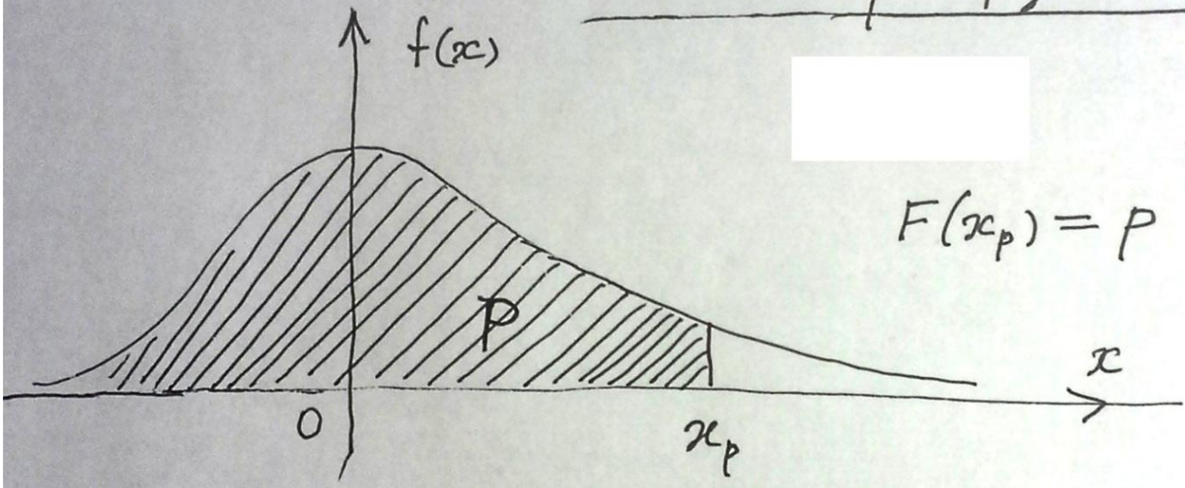
$$F(k_1, k_2) = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}$$

Функция плотности распределения Фишера

$$f_F(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \left(\dfrac{k_1}{k_2}\right)^{k_1/2} \dfrac{x^{\frac{k_1}{2}-1}}{\left(1+\frac{k_1}{k_2}\right)^{(k_1+k_2)/2}} & \end{cases}$$



Квантиль распределения



$$F(x_p) = p$$

$$F(x_p) = P\{X < x_p\}$$

Решения задач

8.529 $\quad P\{Z=k\} = (k-1)p^2 q^{k-2}, \quad k = 2, 3, \ldots$

$$p_k = q^{k-1} p$$

X

1	2	3	4	5	...
p	qp	q^2p	q^3p	q^4p	...

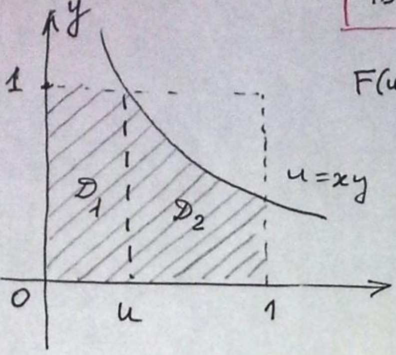
Y

1	2	3	4	5	...
p	qp	q^2p	q^3p	q^4p	...

X + Y

2	3	4
p^2	$2qp^2$	$3q^2p^2$

18.518 б)



$$F(u)=P\{U<u\}=P\{U<XY\}$$

Очевидно $F(u)=0$, при $u<0$

и $F(u)=1$, при $u>1$.

$$F(u)=\iint\limits_{D_1} f(x,y)\,dx\,dy+\iint\limits_{D_2} f(x,y)\,dx\,dy=$$

$$=\int_0^u dx\int_0^1 dy\cdot(x+y)+\int_u^1 dx\int_0^{u/x^2} dy\,(1+y)=$$

$$=\int_0^u \left(xy+\frac{y^2}{2}\right)\Big|_0^1 dx+\int_u^1\left(xy+\frac{y^2}{2}\right)\Big|_0^{u/x}=$$

$$=\int_0^u\left(x+\frac{1}{2}\right)dx+\int_u^1\left(x\cdot\frac{u}{x}+\frac{1}{2}\,\frac{u^2}{x^2}\right)dx=$$

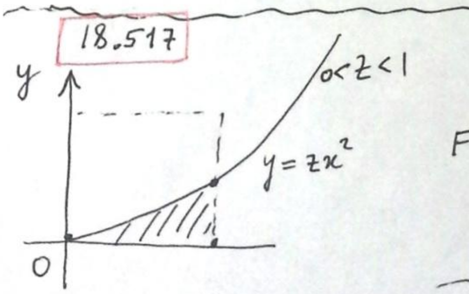
$$=\left(\frac{x^2}{2}+\frac{x}{2}\right)\Big|_0^u+\left(ux-\frac{u^2}{2x}\right)\Big|_u^1=2u-u^2$$

$$F(u)=\begin{cases}0, & u<0\\ 2u-u^2, & 0\le u<1\\ 1, & u\geqslant 1\end{cases},$$

соответственно

$$f(u)=F'(u)=\begin{cases}0, & u<0\\ 2(1-u), & 0\le u<1\\ 0, & u\geqslant 1\end{cases}$$

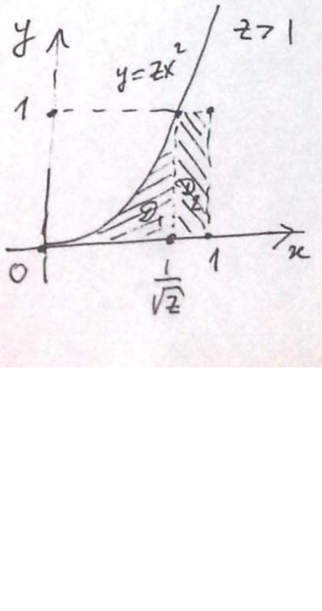
18.517



$$Z=\frac{Y}{X^2},\quad F(z)=P\{Z<z\}$$

$$F(z)=\int_0^1 zx^2\,dx=z\,\frac{x^3}{3}\Big|_0^1=\frac{z}{3}$$

$$f(z)=F'(z)=\frac{1}{3},\quad 0<z<1$$



$$F(z)=\int_0^{1/\sqrt{z}} zx^2\,dx+\int_{1/\sqrt{z}}^1 dx=$$

$$=1-\frac{2}{3}\,\frac{1}{\sqrt{z}},\quad z>1$$

$$f(z)=F'(z)=\frac{1}{3z^{3/2}},\quad z>1$$

при $z<0$: $F(z)=0 \Rightarrow f(z)=0$