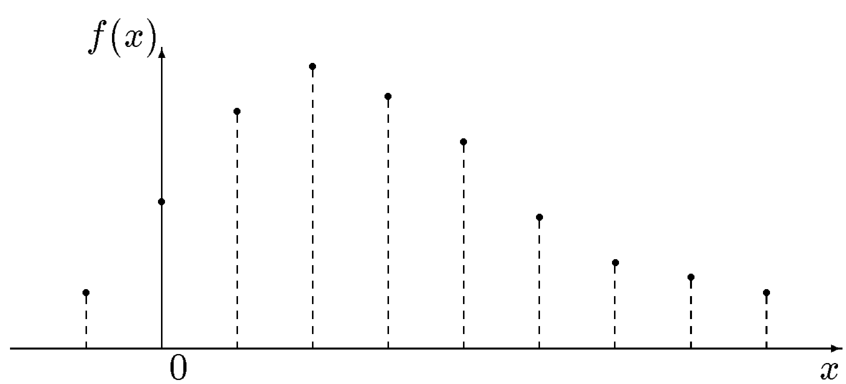
§6. Характеристики распределения случайной величины. Классификация распределений

Мы построили шесть вероятностных моделей, и если пред нами стоит задача их классификаци, то первая очевидная особенность, которой обладает каждое из распределений соответствующей случайной величины, это – непрерывность или разрывность функции распределения. Полученные семейства распределений можно разбить на два класса – *дискретный* и *непрерывный.*

Гипергеометрическое GG(N,M,n), биномиальное B(n,p), пуассоновское P(λ) и геометрическое Geo(p) распределения принадлежат к дикретному классу. При выводе этих распределений мы вполне могли бы ограничиться техникой элементарной теории вероятностей, поскольку пространства элементарных исходов (значений случайной величины X) состояли из конечного или счетного числа точек, и функции плотности $f(x\,|\,\theta)$ в области их ненулевых значений определяли вероятности каждого элементарного исхода $X = x$. Графическое изображение $f(x) = f(x\,|\,\theta)$ как функции x при каждом фиксированном θ позволяет наиболее полно представить картину общего распределения вероятностей и, одновременно, вызывает некоторые ассоциации с "нагруженным стержнем", а также стремление характеризовать распределение масс по стержню такими механическими характеристиками, как центр тяжести, момент инерции, асимметрия и эксцесс в распределении масс и пр.



Прибегая к такой "механической" интерпретации распределения вероятностей, мы соотносим вероятность события $X \in B$ при любом $B \in \mathcal{B}$ с массой участка стержня B и вычисляем величину этой массы по формуле

$$P(B) = \sum_{x \in B} f(x).$$

Центр тяжести нагруженного стержня называется *средним значением* случайной величины X, обозначается $\mathbf{E}X$ и вычисляется как

$$\mathbf{E}X = \sum_{x \in \mathrm{R}} x f(x).$$

Момент инерции относительно точки $\mu = \mathbf{E}X$, равный

$$\mathbf{D}X = \sum_{x \mathrm{R}} (x - \mu)^2 f(x),$$

характеризует меру разброса (удаленности) отдельных точек нагружения от центра масс, и поэтому в теории вероятностей называется *дисперсией* случайной величины X. Кроме стандартного обозначения $\mathbf{D}X$, за величиной дисперсии закреплен символ σ^2, в то время как квадратный корень из дисперсии $\sigma = \sqrt{\mathbf{D}X}$ называется *стандартным отклонением* X.

Несомненный практический интерес представляет также точка достижения максимума функции $f(x)$, как наиболее вероятного значения X. Эта точка называется *модой* распределения X, и как-то так сложилось, что стандартного, наиболее распространенного обозначения у этой характеристики нет, разве лишь $\mathrm{mod}(X)$.

Мы не будем торопиться с введением других характеристик распределения X, а также иллюстрировать вычисления $\mathbf{E}X$, $\mathbf{D}X$ и $\mathrm{mod}(X)$ на конкретных распределениях, и сначала попытаемся ввести аналоги этих характеристик для случайных величин с непрерывной функцией распределения.

К классу непрерывных распределений принадлежат равномерное U(a,b) и показательное $\mathrm{E}(\theta)$ распределения. При построении этих вероятностных моделей функция распределения играла определяющую роль и теорема 4.1 использовалась по существу.

Графическое изображение непрерывной функции распределения вряд ли стоит рассматривать как столь же наглядную иллюстрацию распределения вероятностей, как, например, график функции плотности (функции скачков) распределения дискретного типа. Это замечание в равной степени относится как к дискретному, так и непрерывному классу распределений. Графики возрастающих функций с областью значений в интервале $[\,0;\,1\,]$ так похожи друг на друга, что их главная примечательность – точки перегиба – "на глаз" определяются только при высоких художественных достоинствах графического изображения. Другое дело – производная функции, где эти точки перегиба превращаются в точки экстремума. С другой стороны, производная функции распределения в непрерывном случае, так же как и функция скачков дискретного распределения, допускает механическую интерпретацию функции плотности единичной массы, "размазанной" по бесконечному стержню, и в рамках этой интерпретации мы снова можем рассматривать такие характеристики, как центр тяжести, момент инерции и тому подобное.

Итак, определим *функцию плотности* непрерывного распределения $F(x)$ как производную $f(x) = dF(x)/dx$, которая в нашем случае определяется почти всюду по мере Лебега, что, как будет в дальнейшем, вполне достаточно для вычисления характеристик непрерывного распределения. Так, для равномерного распределения $f(x) = f(x\,|\,\theta) = 0$ равна нулю вне сегмента $[\,a;\,b\,]$ и $f(x\,|\,\theta) = (b-a)^{-1}$, то есть постоянна на этом сегменте. В случае показательного распределения $f(x\,|\,\theta) = 0$ при $x < 0$,

$$f(x\,|\,\theta) = \frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\},$$

если $x \geq 0$, и отнесение точки $x = 0$ к области нулевых значений функции f очевидно не изменит значений интегральных характеристик распределения; аналогичное заключение можно сделать и относительно концевых точек a и b равномерного распределения $\mathrm{U}(\mathrm{a},\mathrm{b})$.

Функция распределения из непрерывного класса выражается через свою функцию плотности в виде

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

а вероятность „попадания" X в некоторое произвольное борелевское множество B (вероятность события B) записывается как

$$P(X \in B) = \int_B f(x)dx = \int_{\mathrm{R}} \mathbf{I}_B(x)f(x)dx,$$

где $\mathbf{I}_B(x)$ – индикаторная функция множества B. Естественно, в силу явной нерегулярности (разрывности и прочих пакостей) подынтегральных функций интегралы в этих формулах следует рассматривать как интегралы Лебега по лебеговой мере dx на борелевской прямой $(\mathrm{R}, \mathcal{B})$.

Центр тяжести стержня с непрерывным распределением масс, которое определяется функцией плотности $f(x)$, вычисляется по известной формуле

$$\mu = \mathbf{E}X = \int\limits_{-\infty}^{\infty} xf(x)dx$$

и называется, как и в дискретном случае, *средним значением* случайной величины X. Точно так же момент инерции

$$\sigma^2 = \mathbf{D}X = \int\limits_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

называется *дисперсией* X, а σ – *стандартным отклонением.* Наконец, точка достижения максимума функции плотности:

$$\mathrm{mod}(X) = \arg\max_{x\in\mathrm{R}} f(x) \ -$$

модой распределения X. Окрестность точки $\mathrm{mod}(X)$ обладает наибольшей концентрацией вероятностной массы.

Лекция 9

Естественно, рассмотрев два основных класса распределений, мы могли бы теперь продолжить изучение характеристик распределений каждого типа, но возникает естественный вопрос, а существуют ли смешанные дискретно-непрерывные распределения или вообще распределения, не принадлежащие к изученным классам, и как тогда вычислять их средние значения и дисперсии?

Что касается дикретно-непрерывных распределений, то о существовании и практической ценности таких распределений указывает следующая вероятностная модель теории надежности. Предположим, что предприятие выпускает изделия с показательным распределением долговечности, но в силу специфических дефектов производства каждое изделие с некоторой вероятностью p может быть "мертворожденным", то есть отказать при его "включении". В таком случае функция распределения долговечности в области $x > 0$ имеет вид (используется формула полной вероятности) $F(x) = p + (1-p)(1 - \exp\{-x/\theta\})$, а средний срок службы

$$\mathbf{E}X = 0 \cdot p + (1-p)\theta^{-1}\int_0^\infty x \exp\{-x/\theta\}\,dx = (1-p)\theta\ -$$

опять новая формула для вычисления характеристик распределения X!

Дальше – больше, оказывается существует еще один тип распределений, вычисление характеристик которого вообще немыслимо вне рамок теории интеграла Лебега. Помните, мы говорили с вами о связи между схемой испытаний Бернулли с вероятностью успешного испытания $p = 1/2$ и равномерным распределением на отрезке $[\,0,\,1\,]$? Оказывается, если вероятность успеха $p \neq 1/2$, то двоичная дробь, составленная из реализаций индикаторов успеха, представляет результат наблюдения случайной величины с весьма загадочной функцией распределения. Во-первых, эта функция почти всюду постоянна – производная от нее почти всюду по мере Лебега на $(\mathrm{R},\,\mathcal{B})$ равна нулю. Тем не менее эта функция возрастает, непрерывна(!), но точки ее роста составляют счетное множество, имеющее, естественно, нулевую лебегову меру. Соответствующая этой функции распределения вероятностная мера P на борелевской прямой сингулярна относительно меры Лебега: если множество $B \in \mathcal{B}$ имеет нулевую лебегову меру, то отсюда не следует, что $P(B) = 0$.

Распределения такого вида, имеющие непрерывную функцию распределения, но сингулярные по отношению к мере Лебега, составляют класс *сингулярных* распределений. Легко понять, что явная запись таких распределений вряд ли возможна. В нашем примере с построением реализаций случайной величины X с помощью схемы Бернулли для функции распределения X составляется некоторое операторное уравнение, и если мы хотим рассчитать вероятности попадания X в интервалы на прямой, то придется использовать численные методы решения таких уравнений. Конечно, мы не будем заниматься сейчас выводом уравнения, определяющего сингулярную функцию распределения – это слишком трудоемкая задача, для решения которой у нас нет времени.

Итак, мы рассмотрели три типа распределений: дискретный, непрерывный и сингулярный. Удивительно то, что других типов не существует, о чем свидетельствует знаменитая

Теорема Лебега. *Любая функция распределения представима в виде суммы трех неотрицательных, неубывающих функций, одна из которых абсолютно непрерывна и имеет неотрицательную производную на множестве положительной лебеговой меры; вторая является ступенчатой и обладает не более чем счетным множеством точек разрыва (скачков); третья непрерывна, но имеет не более чем счетное множество точек роста.*

Доказательство этой теоремы выходит из рамок нашего общего курса теории вероятностей. В не столь отдаленные времена, когда в университете занимались преподаванием фундаментальных наук, а не обучением примитивному ремеслу, теорема Лебега доказывалась в общем курсе математического анализа. Из теоремы Лебега вытекает, что в чистом виде существует только три типа распределений, из которых два (непрерывный и дискретный) нам знакомы, а третий – сингулярный – загадочен, и мы пока не в состоянии представить себе, каким образом вычислять интеграл Лебега,

$$\mathbf{E}X = \int\limits_{\mathrm{R}} x dP(x),$$

определяющий среднее значение случайной величины X с сингулярным распределением вероятностей $P(B)$, $B \in \mathcal{B}$.

Спешу обрадовать вас, что мы не будем рассматривать сингулярные вероятностные модели. Тем не менее существует весьма общий подход к определению *функции плотности* для любого, в том числе и смешанного, типов распределений, опираясь на который можно предложить некоторый общий метод определения и вычисления характеристик распределений различных типов. Этот подход указывает следующая, не мене знаменитая, чем теорема Лебега,

Теорема Радона–Никодима. *Пусть на борелевской прямой* $(\mathrm{R}, \mathcal{B})$ *заданы вероятность* P *и сигма-конечная мера* μ*, причем* P *абсолютно непрерывна относительно* μ*, то есть* $\mu(B) = 0$ *влечет* $P(B) = 0$*. Тогда для почти всех по мере* μ *точек* $x \in \mathrm{R}$ *существует такая единственная функция* $f(x)$*, что*

$$P(B) = \int_B f(x) d\mu(x), \quad \forall B \in \mathcal{B}. \tag{1}$$

Эта теорема, доказательство которой мы также опускаем (и не потому, что времени нет, а просто – знаний не хватает), позволяет ввести одно из центральных понятий теории вероятностей, постоянно используемое при построении вероятностных моделей.

Определение 6.1. Функция $f(x)$, определяемая соотношением (1) для почти всех по мере μ точек $x \in \mathrm{R}$, называется *функцией плотности* распределения вероятностей P по мере μ. Эта функция называется также *производной Радона–Никодима* меры P по мере μ, и имеет место символическая запись $f(x) = dP/d\mu$.

В рамках этого определения, введенная выше функция плотности непрерывного распределения есть производная Радона–Никодима вероятности P по мере Лебега $d\mu = dx$ на борелевской прямой. Так как вероятность P в соответствии с теоремой 4.1 определялась с помощью функции распределения $F(x)$, то мы использовали тот вариант производной Радона–Никодима, который совпадает с обычной производной функции $F(x)$, доопределяя эту функцию в точках, где производная не существует таким образом, чтобы не возникали дополнительные разрывы. Что же касается дискретного случая, то здесь мы использовали производную Радона–Никодима по *считающей мере* μ : для любого $B \in \mathcal{B}$ мера $\mu(B)$ равна количеству точек с целочисленными координатами, которые принадлежат B. Например, борелевское множество $B = [-2.5;\ 5]$ содержит восемь точек с целочисленными координатами $-2, -1, 0, \ldots 5$, и поэтому $\mu(B) = 8$. В "дробных" точках $x \in \mathrm{R}$ мы полагали $f(x) = 0$, хотя могли бы выбирать любые другие значения при вычислении вероятностей по формуле (1). Дело в том, что при интегрировании по дискретной считающей мере интеграл Лебега от любой функции превращается в сумму значений этой функции в целочисленных точках, и (1) принимает известный нам из элементарной теории вероятностей вид

$$P(B) = \sum_{x \in B} P(X = x) = \sum_{x \in B} f(x).$$

Теперь мы обладаем общим подходом к определению характеристик распределения случайной величины X. Значительная часть из них определяется через интеграл Лебега по мере (вероятности) P от специально подобранных функций.

Определение 6.2. Пусть X – случайная величина с распределением P и $f(x)$ – функция плотности P по сигма-конечной мере μ. *Математическим ожиданием* любого измеримого отображения $g(X)$ борелевской прямой в себя (измеримой функции от случайной величины X) называется интеграл Лебега

$$\mathbf{E}g(X) = \int_{\mathrm{R}} g(x)dP(x) = \int_{\mathrm{R}} g(x)f(x)d\mu(x).$$

В частности, *иатематическое ожидание случайной величины* X вычисляется по формуле

$$\mathbf{E}X = \int_{\mathrm{R}} xdP(x) = \int_{\mathrm{R}} xf(x)d\mu(x).$$

З а м е ч а н и е. В отечественной литературе по теории вероятностей (например, в учебнике А.А.Боровкова “Теория вероятностей”) математическое ожидание обозначается латинской буквой **M** а не **E**.

Моментные характеристики распределения случайной величины. Математическое ожидание функции $g(X) = (X - a)^k$ от случайной величины X, где k принимает только целочисленные значения $1, 2, \ldots$, называется *моментом k-го порядка случайной величины X относительно точки a.* Если $a = 0$, то $\alpha_k = \mathbf{E}X^k$ называется просто *моментом k-го порядка* случайной величины X, а если $a = \mathbf{E}X$ $(= \alpha_1)$, то момент $\mu_k = \mathbf{E}(X - \mathbf{EX})^k$ называется *центральным моментом k-го порядка.* Иногда, во избежание недоразумений, моменты α_k называются *нецентральными моментами.* Первый нецентральный момент $\alpha_1 = \mathbf{E}X$ называется *средним значением* или *математическим ожиданием* случайной величины X и обозначается обычно буквой μ. Второй центральный момент $\mu_2 = \mathbf{E}(X - \mu)^2$ называется *дисперсией* случайной величины X и обозначается или буквой σ^2, или вводится оператор $\mathbf{D}X$. Напомним, что квадратный корень из дисперсии: $\sigma = \sqrt{\mathbf{D}X}$ мы договорились называть *стандартным отклонением* X. Поскольку σ имеет ту же размерность, что и наблюдаемая случайная величина X, то в практических приложениях в качестве меры “разброса” вероятностей используется обычно стандартное отклонение σ, а не дисперсия σ^2. Для среднего и дисперсии X справедливо

Предложение 6.1. *Среднее значение* $\mathbf{E}X$ *и дисперсия* $\mathbf{D}X$ *обладают следующими свойствами:*

1^0. $\mathbf{E}(aX+b)=a\mathbf{E}X+b$ для любых постоянных $a,b\in\mathrm{R}$,

2^0. $\mathbf{D}(aX+b)=a^2\mathbf{D}X$ для любых постоянных $a,b\in\mathrm{R}$, то есть дисперсия инвариантна относительно сдвигов случайной величины X на постоянную величину;

3^0. $\mathbf{D}X=\mathbf{E}X^2-(\mathbf{E}X)^2=\alpha_2-\mu^2$,

4^0. $\inf\limits_{a\in\mathrm{R}}\mathbf{E}(X-a)^2=\mathbf{D}X$, то есть $\arg\inf\limits_{a\in\mathrm{R}}\mathbf{E}(X-a)^2=\mathbf{E}X$.

Доказательство.

1^0. Данное утверждение есть простая констатация известного свойства линейности интеграла Лебега.

2^0. $\mathbf{D}(aX+b)=\mathbf{E}(aX+b-a\mu-b)^2=a^2\mathbf{E}(X-\mu)^2=a^2\mathbf{D}X$.

3^0. $\mathbf{D}X=\mathbf{E}(X-\mathbf{E}X)^2=\mathbf{E}\left(X^2-2X\mathbf{E}X+(\mathbf{E}X)^2\right)=$
$\mathbf{E}X^2-2\mathbf{E}X\cdot\mathbf{E}X+(\mathbf{EX})^2=\mathbf{E}X^2-(\mathbf{E}X)^2$.

4^0. $\mathbf{E}(X-a)^2=\mathbf{E}\left((X-\mu)-(a-\mu)\right)^2=$
$\mathbf{E}\left((X-\mu)^2-2(a-\mu)(X-\mu)+(a-\mu)^2\right)=\mathbf{E}(X-\mu)^2-2(a-\mu)\mathbf{E}(X-\mu)+(a-\mu)^2=\mathbf{D}X+(a-\mu)^2\geq\mathbf{D}X$, причем равенство достигается тогда и только тогда, когда $a=\mu=\mathbf{E}X$.

С моментами случайной величины X связаны две замечательные *характеристики формы* распределения X :

коэффициент асимметрии $\gamma_1=\mu_3/\sigma^3$, и

коэффициент эксцесса $\gamma_2=\mu_4/\sigma^4-3$.

Легко заметить по аналогии с доказательством пункта 2^0 предыдущего предложения, что γ_1 и γ_2 *инвариантны относительно линейных преобразований случайных величин*, то есть X и $aX+b$ имеют одинаковые коэффициенты асимметрии и эксцесса при любых постоянных a и b.

Как и выше, мы будем называть *модой* распределения случайной величины X любую точку $\mathrm{mod}(X)$ достижения локального максимума у функции плотности $f(x)$. Если мода единственна, то говорят, что распределение X *унимодально.* Когда график унимодальной кривой плотности имеет „длинный хвост" справа от моды (см. рисунок на следующей странице), то в выражении μ_3 кубы положительных отклонений перевесят отрицательные кубы, и коэффициент асимметрии γ_1 будет положителен. Если же мода „свалена" вправо (длинный хвост слева от моды), то $\gamma_1 < 0$. Распределения с симметричной функцией плотности, как, например, биномиальное с $p = 1/2$ или равномерное U(a,b), обладают нулевой асимметрией: $\gamma_1 = 0$.

Что же касается коэффициента эксцесса γ_2, то его подлинный смысл мы поймем после знакомства в следующем параграфе с *нормальным распределением* на борелевской прямой, а пока только отметим, что положительный эксцесс говорит об излишней „пикообразности" – вытянутости вверх кривой плотности, в то время как отрицательное значение γ_2 указывает на более плоский характер вершины кривой плотности.

Лекция 10

Прежде чем перейти к примерам по вычислению моментных характеристик случайных величин, следует обратить на внимание на то, что в рассмотренных нами вероятностных моделях существуют довольно крупные элементы, имеющие нулевую вероятность, например, во всех моделях $P(X \in (-\infty, 0)) = 0$. В связи с этим вводится понятие *носителя* распределения случайной величины, как замыкания множества $\{x \in \mathrm{R} : \ f(x) > 0\}$. Такое определение носителя не является достаточно общим и связано с мерой μ, по которой вычисляется плотность $f(x)$, но поскольку мы договорились рассматривать только дискретные и непрерывные распределения (μ – считающая мера или мера Лебега), то такое определение вполне работоспособно и позволяет легко найти носитель любого из шести известных нам распределений. Носитель распределения будет обозначаться рукописной буквой $\mathcal{X}$. Нетрудно понять, что при вычислении моментных и прочих интегральных характеристик распределения из области интегрирования можно убрать все точки, не принадлежащие $\mathcal{X}$, и при этом величина характеристики останется неизменной.

Пример 6.1 (*биномиальное распределение* $\mathrm{B}(n,\ p)$). Носитель этого распределения $\mathcal{X} = \{0, 1, \ldots, n\}$. Для вычисления первых двух моментов биномиального распределения воспользуемся методом "дифференцирования по параметру" и формулой бинома Ньютона:

$$\sum_{k=0}^{n} C_n^k a^k b^{n-k} = (a+b)^n.$$

По определению среднего значения

$$\mu = \mathbf{E}X = \sum_{k=0}^{n} k C_n^k p^k (1-p)^{n-k} = p\left[\frac{d}{dx}\sum_{k=0}^{n} C_n^k x^k (1-p)^{n-k}\right]_{x=p} =$$

$$p\left[\frac{d}{dx}(x+1-p)^n\right]_{x=p} = pn(x+1-p)^{n-1}\ |_{x=p} = np.$$

Второй момент

$$\alpha_2 = \mathbf{E}X^2 = \sum_{k=0}^{n} k^2 C_n^k p^k (1-p)^{n-k} = p\left[\frac{d}{dx}x\frac{d}{dx}\sum_{k=0}^{n} C_n^k x^k (1-p)^{n-k}\right]_{x=p} =$$

$$p\left[\frac{d}{dx}x\frac{d}{dx}(x+1-p)^n\right]_{x=p} = p\left[\frac{d}{dx}xn(x+1-p)^{n-1}\right]_{x=p} =$$

$$np\left[(x+1-p)^{n-1} + x(n-1)(x+1-p)^{n-2}\right]_{x=p} = np(1-p) + (np)^2,$$

откуда дисперсия биномиального распределения $\sigma^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2 = np(1-p)$.

С помощью аналогичных, но более утомительных выкладок можно найти третий и четвертый моменты, а также коэффициенты асимметрии и эксцесса

$$\gamma_1 = \frac{1-2p}{\sqrt{np(1-p)}},\ \ \gamma_2 = \frac{1-6p(1-p)}{np(1-p)}.$$

Следовательно, биномиальное распределение “свалено” влево (длинный хвост справа) при $p < 1/2$, симметрично, как нам было известно ранее, при $p = 1/2$ и “свалено” вправо при $p > 1/2$. Коэффициент эксцесса положителен в области $p(1-p) < 1/6$, а наибольшее по абсолютной величине отрицательное значение $\gamma_2 = -2/n$, когда $p = 1/2$.

Мода $\mathrm{B}(n,p)$ определяется как целочисленное x, при котором происходит смена неравенства $f(x \,|\, n,p) < f(x+1 \,|\, n,p)$ на обратное. Нетрудно убедиться, что это неравенство эквивалентно $x+1 < p(n+1)$, так что $\mathrm{mod}(X)$ определяется через сравнение значений $f(x \,|\, n,p)$ при целых $x \geq 0$, ближайших к $p(n+1)$.

Пример 6.2 (*распределение Пуассона* $\mathrm{P}(\lambda)$). Носитель распределения $\mathcal{X} = \{0, 1, \ldots, \infty\}$ – точка $x = \infty$ должна быть включена в носитель по требованию замыкания множества вероятности единица. Моментные характеристики пуассоновского распределения можно рассчитать, используя тот же метод дифференцирования по параметру, но проще, вспомнив, что $\mathrm{P}(\lambda)$ есть предел B(n,p) при $n \to \infty$, $p \to 0$ и $np = \lambda$, перейти к этому пределу в моментных характеристиках биномиального распределения. В результате получаем

$$\mathbf{E}X = \mathbf{D}X = \lambda, \ \gamma_1 = \lambda^{-1/2}, \ \gamma_2 = \lambda^{-1},$$

а $\mathrm{mod}(X) = [\,\lambda\,]$, поскольку асимметрия $\mathrm{P}(\lambda)$ всегда положительна и график $f(x \,|\, \lambda)$ “свален” влево.

Следует обратить особое внимание на то, что у распределения Пуассона дисперсия совпадает со средним значением: $\mu = \sigma^2 = \lambda$.

Пример 6.3 (*равномерное распределение* $\mathrm{U}(a,\, b)$). Носитель распределения $\mathcal{X} = [\,a;\, b\,]$. Модой распределения является любая точка интервала $(a,\, b)$, поскольку плотность $f(x) = (b-a)^{-1}$ постоянна на этом интервале.

Нетрудно убедиться, что если случайная величина X имеет распределение $\mathrm{U}(0,\,1)$, то $Y = (b-a)X + a,\ b > a$, распределена как $\mathrm{U}(a,\, b)$. Это проистекает из-за следующего соотношения между функциями распределения случайных величин: $P(Y < x) = P((b-a)X + a < x) = P(X < (x-a)/(b-a)) = (x-a)/(b-a)$. В силу этого для вычисления моментных характеристик $\mathrm{U}(a,\, b)$ достаточно найти соответствующие характеристики $\mathrm{U}(0,1)$ и затем воспользоваться предложением 6.1.

Для распределения $\mathrm{U}(0,1)$ имеем

$$\mu = \mathbf{E}X = \int_0^1 x dx = 1/2, \quad \alpha_2 = \int_0^1 x^2 dx = 1/3,$$

откуда дисперсия $\sigma^2 = 1/3 - 1/4 = 1/12$. Следовательно, для распределения $\mathrm{U}(a, b)$ (см. предложение 6.1) $\mu = a + (b-a)/2$, $\sigma^2 = (b-a)^2/12$.

Симметричное равномерное распределение $\mathrm{U}(a, b)$ имеет нулевой коэффициент асимметрии, в то время как коэффициент эксцесса γ_2 отрицателен (не будем заниматься его вычислением).

Пример 6.4 (*показательное распределение* $\mathrm{E}(\theta)$). Носитель распределения $\mathcal{X} = [\,0, \infty\,]$ – расширенная положительная часть прямой R. Наибольшее значение плотности $f(x) = \theta^{-1} \exp\{-x/\theta\}$ достигается в точке $x = 0$, поэтому $\mathrm{mod}(X)$=0.

Моменты показательного распределения

$$\alpha_k = \theta^{-1} \int_0^\infty x^k \exp\{-x/\theta\} dx = \theta^k \int_0^\infty x^k e^{-x} dx = \Gamma(k+1)\theta^k = k!\theta^k,$$

откуда $\mu = \theta, \quad \sigma^2 = \theta^2$ и стандартное отклонение $\sigma = \theta$ совпадает со средним значением.

Естественно, моментные характеристики далеко не универсальны, и можно привести примеры распределений, у которых существует ограниченное количество моментов, или не существует даже среднего значения. Мы приведем два из таких распределений, одно из которых может представлять некоторый практический интерес, а другое будет использоваться для иллюстраций различных паталогий в теории статистического вывода; оба распределения заносятся в каталог вероятностных моделей.

Распределение Парето $\mathrm{Par}(a, \alpha)$. Налоговые органы обычно интересуются распределением годовых доходов тех лиц, годовой доход которых превосходит некоторый предел a, установленный законами о налогообложении. Такого рода распределения иногда считают (к сожалению, без особого “экономического” обоснования) приближенно совпадающими с *распределением Парето*, вся вероятностная масса которого сосредоточена в области $x > a$ (носитель распределения $\mathcal{X} = [\,a, \infty\,]$), и функция распределения на сегменте $\mathcal{X}$ равна

$$F(x) = 1 - \left(\frac{a}{x}\right)^\alpha, \quad x > a, \ \alpha > 0.$$

Это распределение, зависящее от двумерного параметра $\theta = (a, \alpha)$ с параметрическим пространством $\Theta = \mathrm{R}_+ \times \mathrm{R}_+$, принадлежит непрерывному типу; его функция плотности в области $x > a$ равна

$$f(x \mid \theta) = \frac{\alpha}{a}\left(\frac{a}{x}\right)^{\alpha+1}.$$

Момент k-го порядка у распределения Парето существует только при значениях параметра $\alpha > k$, например, неравенство $\alpha > 1$ гарантирует существование среднего значения, которое, как нетрудно подсчитать, равно $\alpha a/(\alpha - 1)$.

Если случайная величина X распределена по закону Парето, то, как легко видеть, $\ln X$ имеет показательное распределение, "сдвинутое вправо" на величину $\ln a$, так как $P(\ln X < x) = P(X < e^x) = F(e^x)$. Это замечание объясняет, почему распределение Парето адекватно описывает распределение наблюдаемых доходов у лиц с невысоким уровнем дохода. Вспомним постулат "отсутствия последействия", приводящий к показательному распределению долговечности: вероятность того, что изделие прослужит промежуток времени, не меньший s, при условии, что оно уже отработало срок t, не зависит от величины t. В основу модели Парето положен тот же принцип, только в мультипликативной, а не в аддитивной, формулировке: *вероятность того, что доход отдельного лица увеличится не меньше, чем в s раз, при условии, что он уже достиг уровня t, не зависит от величины достигнутого уровня.* Иными словами, субъекты наблюдения обладают слишком малыми доходами, чтобы их вложение в какого-либо рода предпринимательскую деятельность приводили к наращиванию денежной массы, или (другая интерпретация) изменчивость дохода за наблюдаемые периоды времени носит случайный характер и не связана с величиной капитала, которым располагают отдельные субъекты. Заметим, что у лиц с высоким уровнем дохода распределение наблюдаемых доходов отлично от распределение Парето. Это так называемое логарифмически нормальное распределение, с которым мы познакомимся несколько позже, освоив новые математические методы построения вероятностных моделей.

Распределение Коши $\mathrm{C}(a, b)$. Орудие с вращающимся лафетом помещается на единичном расстоянии от стены, бесконечно уходящей в обе стороны.

Представим, что стена является действительной прямой R с началом координат в основании перпендикуляра, опущенного из орудия на стену. Ствол орудия размещается параллельно стене с направлением выстрела в сторону отрицательной полуоси, лафет орудия начинает равномерно вращаться по ходу часовой стрелки, и прежде, чем ствол займет первое положение параллельное стене, в случайный момент времени происходит выстрел. Экспериментатора интересует распределение случайной величины X, реализация x которой совпадает с координатой точки попадания снаряда.

Пусть φ – случайная величина, соответствующая величине угла, между перпендикуляром к стене и положением ствола в момент выстрела. Нам будет удобнее измерять φ в пределах $[\,-\pi/2;\,\pi/2\,]$ и трактовать предположение о случайном моменте выстрела в терминах равномерного распределения φ на этом сегменте. Следовательно, функция распределения φ при $-\pi/2 \le x \le \pi/2$ равна $F(x) = (x+\pi/2)\pi^{-1}$. Очевидно, координата точки попадания (см. рисунок) $X = \operatorname{tg}\varphi$, откуда искомая функция распределения

$$F(x) = P(X < x) = P(\operatorname{tg}\varphi < x) = P(\varphi < \operatorname{arctg} x) =$$

$$\frac{1}{\pi}\left(\operatorname{arctg} x + \frac{\pi}{2}\right),\ x \in \mathrm{R},$$

а функция плотности

$$f(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

Сдвиг вправо на параметр a и выбор масштабного параметра b определяет то распределение, которому мы присвоим имя Коши и будем обозначать $\mathrm{C}(a,\,b)$; его функция плотности

$$f(x\,|\,a,b) = \frac{1}{\pi b}\left[1+\left(\frac{x-a}{b}\right)^2\right]^{-1},$$

носителем распределения является расширенная числовая прямая $\mathfrak{X} = \bar{\mathrm{R}} = [-\infty, +\infty]$.

Легко видеть, что распределение Коши не обладает даже конечным средним значением, не говоря о моментах более высокого порядка. Однако это распределение симметрично и имеет ярко выраженную моду, $\mathrm{mod}(X)=a$, которая с успехом заменяет среднее значение, как характеристику положения центра масс. В связи с этим полезно сделать замечание о среднем значении как характеристике положения: оно действительно играет свою роль только в случае симметричных распределений, но при больших абсолютных значениях γ_1 среднее перестает быть полезной характеристикой распределения, в то время как мода "всегда хороша".

Какие же характеристики используются при описании распределений, у которых отсутствуют моменты?

Определение 6.3 Пусть функция распределения $F(x)$ случайной величины X строго возрастает в области всех значений своего аргумента, для которых $0 < F(x) < 1$. Тогда для любого $p \in (0; 1)$ корень $x_p = F^{-1}(p)$ уравнения $F(x) = p$ называется *p-квантилью* распределения X.

В том случае, когда $F(x)$ непрерывна, но не строго монотонна, так что уравнение $F(x) = p$ имеет много решений, в качестве p-квантили обычно берется наибольший или наименьший из корней этого уравнения, и выбор корня определяется существом рассматриваемой вероятностной проблемы. В случае же дискретного распределения это уравнение может вообще не иметь решений, и тогда в качестве p-квантили выбирается то значение x, для которого значение $F(x)$ ближе всего к заданному p.

Квантиль считается характеристикой *положения*, и с этой точки зрения особого внимания заслуживает квантиль $x_{0.5}$, которая разделяет всю вероятностную массу на две одинаковые половинки. Эта квантиль носит название *медианы* распределения и обычно обозначается буквой m. У симметричных распределений (биномиальное с вероятностью успешного испытания $p = 1/2$, равномерное и Коши) медиана совпадает с центром симметрии распределения, а при наличии среднего значения у симметричного распределения медиана $m = \mathbf{E}X$. Если p кратно 0.1, то квантиль называется *децилью*, а если $p = 1/4$

или $3/4$, то – *квартилью.*

С квантилями связаны также несколько характеристик *рассеяния* распределения вероятностей. Очевидно, интервал $(x_{1-p};\, x_p)$ при достаточно близких к единице значениях p накрывает основную часть вероятностной массы, и поэтому разность $x_p - x_{1-p},\ p > 1/2$, служит характеристикой *толерантности* распределения случайной величины X. Если $p = 3/4$, то разность $x_{3/4} - x_{1/4}$ называется *семиинтерквартильной широтой* распределения X.

<u>*Лекция 11*</u>

Мы завершим этот параграф доказательством одного замечательного неравенства, играющего исключительную роль при доказательстве многих теорем (или, как часто говорят, "законов") теории вероятностей. Это неравенство или, в большей степени, следствие из него связывает квантильные и моментные характеристики рассеяния распределения.

Предложение 6.2 (неравенство Чебышева). *Для любой неотрицательной измеримой функции $g(x)$ и любого $\varepsilon > 0$ имеет место неравенство*

$$P(\, g(X) > \varepsilon\,) \leq \frac{\mathbf{E}\, g(X)}{\varepsilon}.$$

Доказательство. Если $\mathbf{E}\, g(X) = +\infty$, то неравенство тривиально. В случае конечного математического ожидания

$$\mathbf{E}\, g(X) = \int_{\mathrm{R}} g(x)dP(x) = \int_{g(x)<\varepsilon} g(x)dP(x) + \int_{g(x)\geq\varepsilon} g(x)dP(x).$$

Если в правой части этого равенства первое слагаемое заменить нулем (оно неотрицательно), а во втором слагаемом под интегралом вместо $g(x)$ подставить его наименьшее значение ε, то получим оценку снизу

$$\mathbf{E}\, g(X) \geq \varepsilon \int_{g(x)>\varepsilon} dP(x) = \varepsilon P(g(X) > \varepsilon),$$

из которой немедленно следует неравенство Чебышева.

Следствие 6.1. *Для любой случайной величины X с конечным средним значением $\mathbf{E}X$ и любого $\varepsilon > 0$ имеет место неравенство*

$$P\left(|X - \mathbf{E}X| > \varepsilon\right) \leq \frac{\mathbf{D}X}{\varepsilon^2}. \tag{2}$$

Д о к а з а т е л ь с т в о. Если дисперсия X не существует (равна бесконечности), то утверждение следствия тривиально. В случае $\mathbf{D}X < \infty$ достаточно заменить событие $|X - \mathbf{E}X| > \varepsilon$ на эквивалентное $|X - \mathbf{E}X|^2 > \varepsilon^2$ и применить неравенство Чебышева.

Доказанное неравенство часто используется на практике для универсальной характеристики толерантности распределений, обладающих конечным средним μ и конечной дисперсией σ^2. Имеется в виду распространенное

Правило трех сигм. *Интервал с концами $\mu \pm \sigma$ содержит приблизительно 90% вероятностной массы распределения X.*

Действительно, если в неравенстве (2) положить $\varepsilon = 3\sigma$, то получим: $P(\mu - 3\sigma \leq X \leq \mu + \sigma) = 1 - P(|X - \mu| > 3\sigma) \geq 8/9 \approx 0.9$.

Так как правило 3σ носит универсальный характер, то оно дает в большинстве случаев слишком грубую оценку толерантности распределения. Например, можно доказать, что для симметричных распределений с конечным третьим моментом μ_3 справедливо *правило* 2σ: интервал с концами $\mu \pm 2\sigma$ содержит 90% вероятностной массы распределения.

В дальнейшем, чтобы не писать длинные названия рассмотренных нами распределений, мы будем указывать распределение X посредством ссылки на символ этого распределения, используя при этом знак эквивалентности, например, $X \sim \mathrm{B}(n, p)$ означает, что X имеет биномиальное распределение.