



ТЕОРИЯ РИСКА

лекции

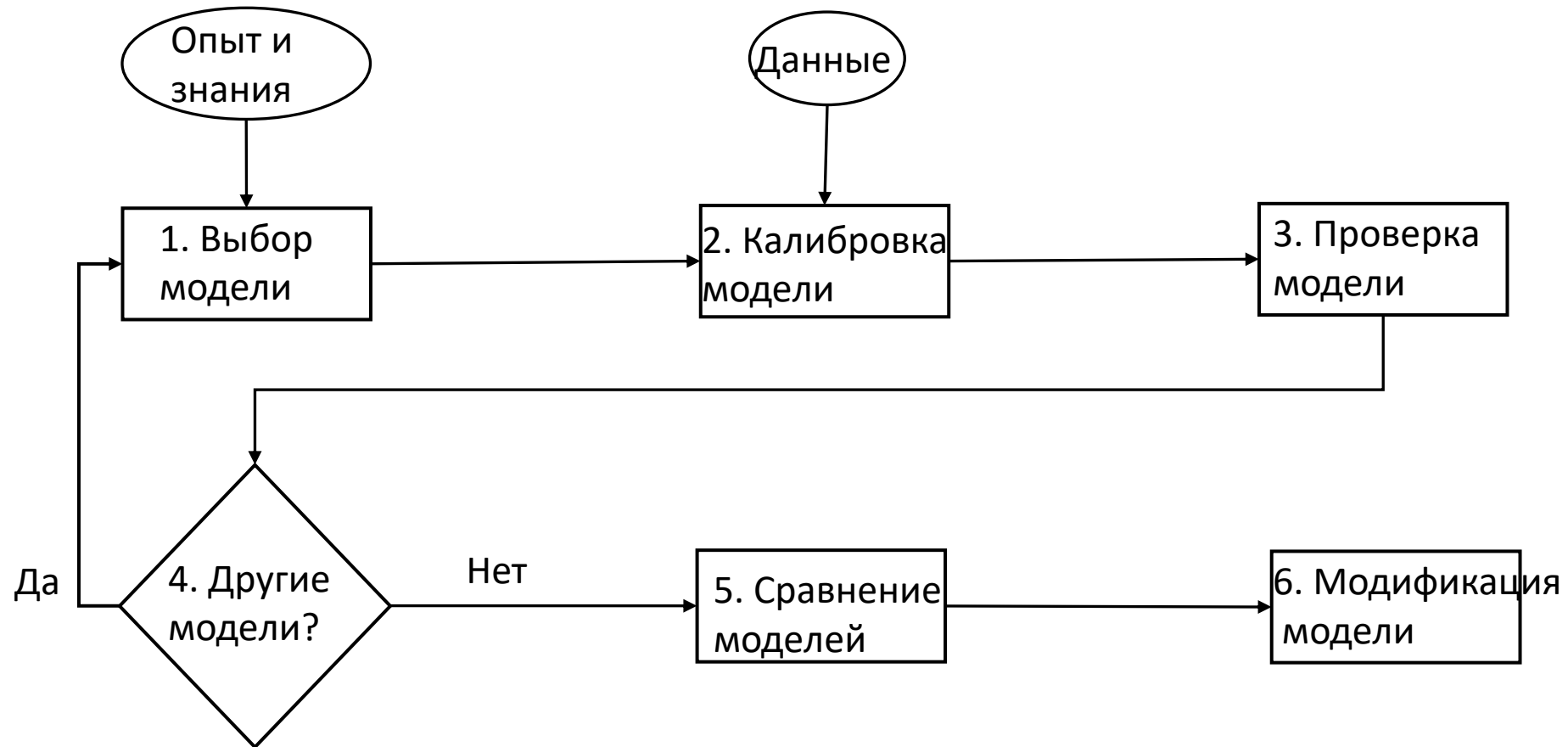
Поспелова Ирина Игоревна



Моделирование

- Модель зависит от целей исследования и особенностей задачи.
- Большинство задач страхования и финансовой математики предполагают построение математической модели с целью прогнозирования или предсказания будущих финансовых потоков.
- Модель – упрощенное математическое описание, основанное на имеющихся знаниях и опыте, учитывающее существующие данные и информацию.
- Наличие и состав данных определяет тип модели, а также набор и калибровку параметров.
- Модель должна обеспечивать баланс между простотой и точностью. Простота определяется, в частности, числом неизвестных параметров; точность оценивается соответствием имеющихся данных модельным результатам.

Процесс моделирования



Процесс моделирования

Шаг 1. На основании опыта и знаний, а также, возможно, формата и природы данных, выбирается одна или несколько подходящих моделей.

Шаг 2. Модель калибруется (подгоняется) под имеющиеся данные.

Шаг 3. Модель проверяется на соответствие данным. Способ проверки зависит от типа модели и типа данных (статистические тесты, критерии качества, ретро-прогноз).

Шаг 4. Рассматриваются другие возможные модели. В особенности это необходимо, если на Шаге 3 модель показала плохое соответствие данным.

Шаг 5. Модели сравниваются, отбирается лучшая.

Шаг 6. Модель адаптируется для применения в будущем. Для этого в модель могут включаться параметры, отражающие экономические, правовые изменения.

Преимущество метода моделирования

- Альтернативой модели является эмпирический подход – принятие решений на основе опыта, наблюдений. Обычно предполагается, что в будущем следует ожидать повторения прошлого, возможно, с поправками на известные тенденции (например, инфляцию).
- **ПРИМЕР 1.**
- Предложим, что 1000 сотрудников различного возраста застрахованы на случай смерти (выплаты могут быть различными, связаны с зарплатой). Наблюдения показывают, что за последние пять лет 14 сотрудников умерли, и в общей сложности выплаты составили 580 000 (с поправкой на инфляцию, поскольку выплата привязана к заработной плате). Определим эмпирическую оценку ожидаемых выплат в следующем году.

Преимущество метода моделирования

- Эмпирическая оценка на следующий год составляет 116 000 (1/5 от 580 000), которые необходимо будет дополнительно скорректировать с учетом роста заработной платы и соответствующих выплат.
- Опасность такой оценки заключается в том, что маловероятно, что опыт последних пяти лет точно повторится в будущем для этого страхового портфеля, поскольку в таких краткосрочных результатах могут наблюдаться значительные колебания.
- Представляется гораздо более разумным построить модель, в данном случае таблицу смертности, которая основана на данных о большом числе людей, а не только 1000 человек в рассматриваемой группе. С помощью такой модели можно не только оценить ожидаемый платеж на следующий год, но и измерить связанный с этим риск, рассчитав стандартное отклонение платежей или, возможно, различные процентиля распределения платежей.

Случайные величины

Модели страхования необходимы для представления потока неизвестных будущих платежей.

«Неизвестный» - неопределенный либо случайный.

Неопределенная величина не предполагает наличия о ней никакой информации, кроме, возможно, множества допустимых значений $\Rightarrow$ модели и методы исследования операций (maxmin), часто приводят к слишком пессимистичной оценке.

Дополнительная информация позволяет либо получить минимаксную оценку (редко), либо перейти к случайной величине, имеющей некоторое, возможно неизвестное, распределение.

Вероятностные модели, как правило, применяется к финансовым потокам.

Неизвестные величины в страховании:

- события (будет ли вообще выплата?)
- момент времени (когда платеж должен быть произведен?)
- размер ущерба (сколько нужно заплатить?)

Распределение случайной величины считаем заданным, вопрос выбора подходящего распределения пока не обсуждаем.

Случайные величины

- Все модели случайных явлений имеют схожие элементы: есть набор возможных исходов. Конкретная реализация определяется исследуемой задачей. Приписывание вероятностей к различным результатам позволяет количественно оценить ожидания и риск их невыполнения. Наименование случайных величин и конкретные характеристики определяются в рамках определенной модели.
- *Явления* – это то, что можно наблюдать.
- *Эксперимент* - это наблюдение за данным явлением при определенных условиях. Результат эксперимента называется *исходом*; *событие* - это набор из одного или нескольких возможных исходов. *Стохастическое явление* - это явление, для которого эксперимент имеет более одного возможного исхода.
- Событие, связанное со стохастическим явлением, называется *случайным*.
- *Вероятность* - это мера возможности наступления события, измеряемая от 0 до 1.
- *Случайная величина* - присваивает числовое значение для каждого возможного исхода.

Иллюстративные модели

Случайные величины, которые могут встречаться в страховой математике

- Возраст на момент смерти при случайно выбранном моменте рождения (**МОДЕЛЬ 1**);
- время до момента смерти с момента приобретения страховки для случайно выбранного застрахованного;
- время от наступления события, приводящего к инвалидности, до выздоровления или смерти для случайно выбранного лица, претендующего на компенсацию;
- время с момента возникновения случайно выбранной претензии до сообщения о ней страховщику; время с момента сообщения о случайно выбранной претензии до ее урегулирования; сумма денег, выплаченная по случайно выбранному страхованию жизни;
- сумма, выплаченная по случайно выбранному иску о нанесении повреждений автомобилю (**МОДЕЛЬ 2**);

Иллюстративные модели

- количество заявлений о телесных повреждениях в результате автомобильной аварии за один год для случайно выбранного застрахованного автомобиля (**МОДЕЛЬ 3**);
- общая сумма исков о врачебной халатности, выплаченная за один год в случайно выбранной больнице (**МОДЕЛЬ 4**);
- время дефолта или предоплаты по случайно выбранному застрахованному ипотечному кредиту, срок действия которого истекает досрочно;
- сумма, выплачиваемая в момент погашения по случайно выбранной высокодоходной облигации;
- значение фондового индекса на определенную дату в будущем.

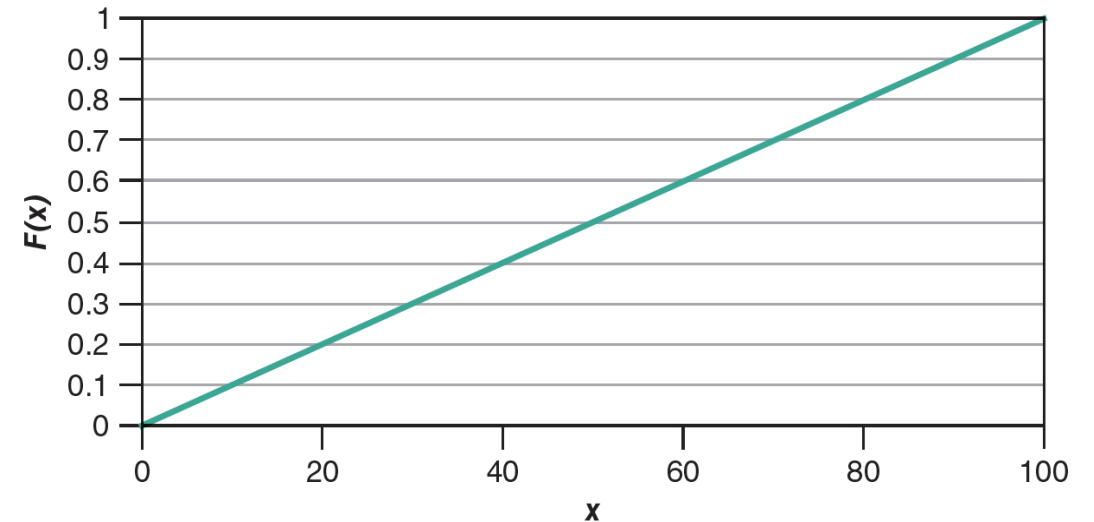
Функция распределения

- **Определение 1.** Функция распределения $F_X(x)$ (или $F(x)$) случайной величины X
$$F_X(x) = P\{X \leq x\}$$
- Свойства:
- $0 \leq F(x) \leq 1 \quad \forall x$
- $F(x)$ не убывает
- $F(x)$ непрерывна справа
- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$

Функция распределения для Модели 1

Случайная величина, отвечающая возрасту смерти, допустимый возраст от 0 до 100:

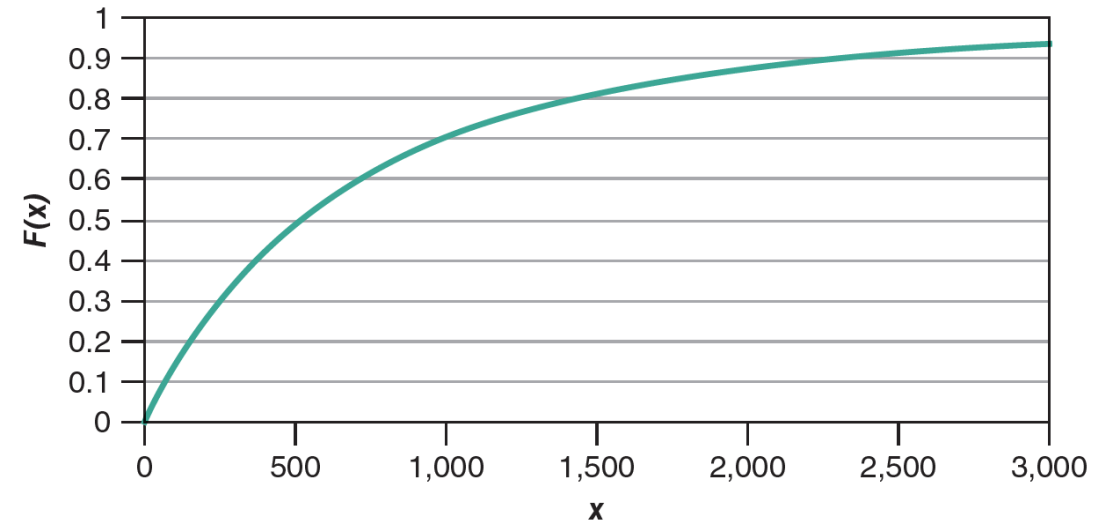
$$F_1(x) = \begin{cases} 0, & x < 0 \\ 0.01x, & 0 \leq x \leq 100 \\ 1, & x \geq 100 \end{cases}$$



Функция распределения для Модели 2

Случайная величина, описывающая количество денег, выплаченных по иску автострахования. Допустимы все положительные значения. Как и в случае смертности, скорее всего, есть верхний предел, однако в модели это требование не обязано выполняться:

$$F_2(x) = \begin{cases} 0, & x < 0 \\ 1 - \left(\frac{2000}{x + 2000} \right)^3, & x \geq 0 \end{cases}$$



Функция распределения для Модели 3

Случайная величина описывает число исков на один полис в год. Вероятности определены для значений 0, 1, 2, 3, 4:

$$F_3(x) = \begin{cases} 0, & x < 0, \\ 0.5, & 0 \leq x < 1, \\ 0.75, & 1 \leq x < 2 \\ 0.87, & 2 \leq x < 3 \\ 0.95, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

Функция распределения для Модели 4

Случайная величина отвечает суммарному количеству денег выплаченных по полису медицинской халатности в год. В большинстве лет ничего не выплачивается, поэтому вероятность нулевой выплаты велика (равна 0.7). Оставшаяся часть, т.е. 0.3, распределена на множестве положительных значений:

$$F_4(x) = \begin{cases} 0, & x < 0 \\ 1 - 0.3e^{-0.00001x}, & x \geq 0 \end{cases}$$

Типы случайных величин

- **Определение 2.** Носителем случайной величины называется множество ее возможных значений N_X : $P\{X \notin N_X\} = 0$.
- **Определение 3.** Случайная величина называется дискретной, если ее носитель содержит не более счетного числа значений.
Случайная величина называется *непрерывной*, если функция распределения непрерывна и дифференцируема всюду за исключением, возможно, счетного числа точек.
Случайная величина называется *смешанной*, если она не является дискретной и является непрерывной всюду, за исключением не менее одного значения и не более чем счетного числа значений.

Случайные величины в Моделях

- **ПРИМЕР 2.** Найдем носитель для каждой из четырех моделей и определим тип случайной величины.

- **Модель 1:**
$$F_1(x) = \begin{cases} 0, & x < 0 \\ 0.01x, & 0 \leq x \leq 100 \\ 1, & x \geq 100 \end{cases}$$

- Функция распределения непрерывна и дифференцируема, кроме точек 0 и 100, так что это непрерывная случайная величина. Носитель включает в себя все значения между 0 и 100. Но сами 0 и 100 могут включаться или нет в зависимости от конкретного смысла случайной величины.

- **Модель 2:**
$$F_2(x) = \begin{cases} 0, & x < 0 \\ 1 - \left(\frac{2000}{x+2000}\right)^3, & x \geq 0 \end{cases}$$

- Функция распределения непрерывна и дифференцируема всюду, кроме 0 (непрерывная случайная величина). Носитель – все положительные действительные числа и, возможно, 0.

Случайные величины в Моделях

- **Модель 3:**
$$F_3(x) = \begin{cases} 0, & x < 0, \\ 0.5, & 0 \leq x < 1, \\ 0.75, & 1 \leq x < 2 \\ 0.87, & 2 \leq x < 3 \\ 0.95, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$
- Вероятности определены только для значений 0, 1, 2, 3, 4 (это и есть носитель), поэтому случайная величина дискретная.
- **Модель 4:**
$$F_4(x) = \begin{cases} 0, & x < 0 \\ 1 - 0.3e^{-0.00001x}, & x \geq 0 \end{cases}$$
- Функция распределения непрерывна везде, кроме 0, где есть скачок. Это смешанное распределение с носителем, состоящим из неотрицательных действительных чисел.

Функция выживания

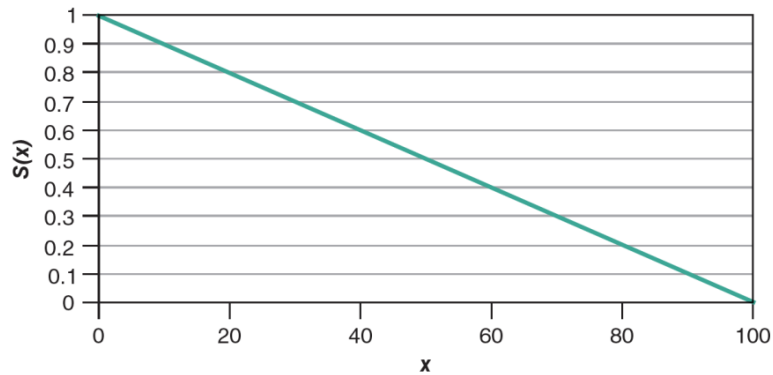
- **Определение 4.** Функцией выживания, обычно обозначаемой через $S_X(x)$ или $S(x)$, для случайной величины X называется дополнительная функция распределения:

$$S_X(x) = P\{X > x\} = 1 - F_X(x)$$

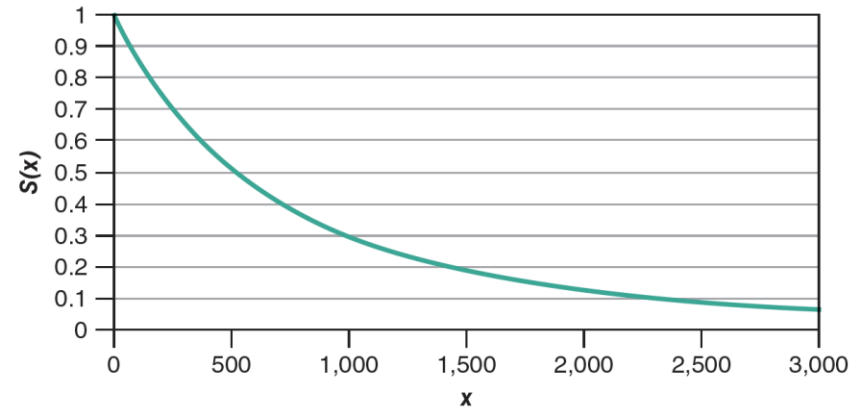
- Свойства:
- $0 \leq S(x) \leq 1 \quad \forall x$
- $S(x)$ не возрастает
- $S(x)$ непрерывна справа
- $\lim_{x \rightarrow -\infty} S(x) = 1, \lim_{x \rightarrow \infty} S(x) = 0$
- Исторически сложилось описывать случайные величины, обозначающие деньги, с помощью функции распределения, а случайные величины, обозначающие время, с помощью функции выживания

Функции выживания для Моделей

$$S_1(x) = 1 - 0.01x, \quad 0 \leq x < 100,$$



$$S_2(x) = \left(\frac{2000}{x+2000} \right)^3, \quad x \geq 0,$$



$$S_3(x) = \begin{cases} 0.5, & 0 \leq x < 1, \\ 0.25, & 1 \leq x < 2, \\ 0.13, & 2 \leq x < 3, \\ 0.05, & 3 \leq x < 4, \\ 0, & x \geq 4 \end{cases}$$

$$S_4(x) = 0.3e^{-0.00001x}, \quad x \geq 0$$

Функция плотности распределения

- **Определение 5.** Функция плотности (вероятности) $f_X(x)$ (или $f(x)$) является производной функции распределения или, что эквивалентно, производной функции выживания с минусом:
$$f(x) = F'(x) = -S'(x).$$

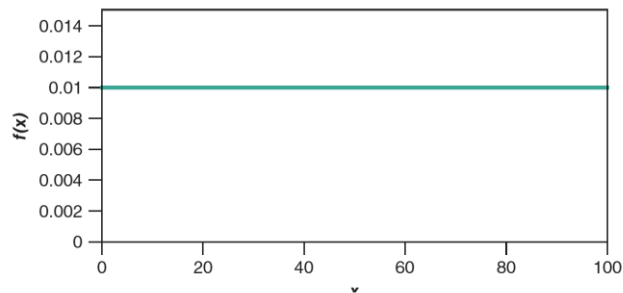
Функция плотности определена только в тех точках, где производная существует.

Функция распределения и функция выживания получаются интегрированием функции плотности по соответствующему интервалу:

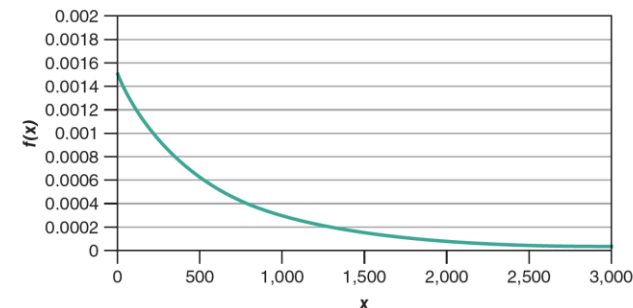
$$P\{a < X \leq b\} = \int_a^b f(x)dx$$
$$F(b) = \int_{-\infty}^b f(x)dx, \quad S(b) = \int_b^{\infty} f(x)dx$$

Функции плотности для моделей

$$f_1(x) = 0.01 \quad 0 < x < 100,$$



$$f_2(x) = \frac{3(2000)^3}{(x + 2000)^4}, x > 0,$$



$$f_3(x) \text{ не определена,}$$

$$f_4(x) = 0.000003e^{-0.00001x}, x > 0$$

Для смешанных распределений (Модель 4) функция плотности не полностью описывает распределение вероятностей

Функция массы вероятности

- **Определение 6.** Функция вероятности или функция массы вероятности обычно обозначается через $p_X(x)$ или $p(x)$ и показывает вероятность в определенной точке (когда она не равна 0):

$$p_X(x) = P\{X = x\}$$

- Определяется для дискретных случайных величин. В этом случае $F(x) = \sum_{y \leq x} p(y)$, $S(x) = \sum_{y > x} p(y)$
Для Моделей: $p_1(x), p_2(x)$ не определены.

$$p_3(x) = \begin{cases} 0.5, x = 0 \\ 0.25, x = 1, \\ 0.12, x = 2, \\ 0.08, x = 3, \\ 0.05, x = 4, \end{cases} \quad p_4(x) = 0.7$$

Так как в Модели 4 распределение смешанное, то функция вероятности описывает только его дискретную часть. Можно было бы представить функцию плотности в виде

$$f_4(x) = \begin{cases} 0.7, x = 0 \\ 0.000003e^{-0.00001x}, x > 0 \end{cases}$$

понимая, что формально это не является ни функцией вероятности, ни плотностью

Уровень риска (hazard rate)

Определение 7. Уровень (интенсивность) риска, также называемый интенсивностью смертности и интенсивностью отказов, обозначается через $h_X(x)$ или $h(x)$:

$$h_X(x) = \frac{f_X(x)}{S_X(x)}$$

Для интенсивности смертности используется обозначение $\mu(x)$, для интенсивности отказов $\lambda(x)$.

Нетрудно видеть, что

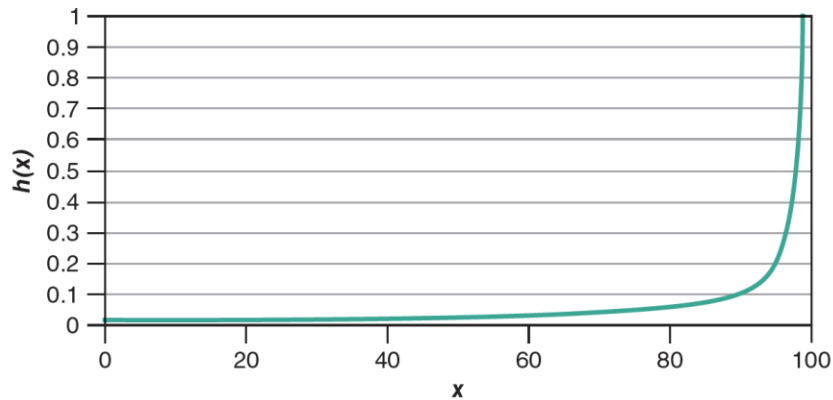
$$h_X(x) = \frac{-S'(x)}{S(x)} = -\frac{d \ln S(x)}{dx}.$$

Отсюда

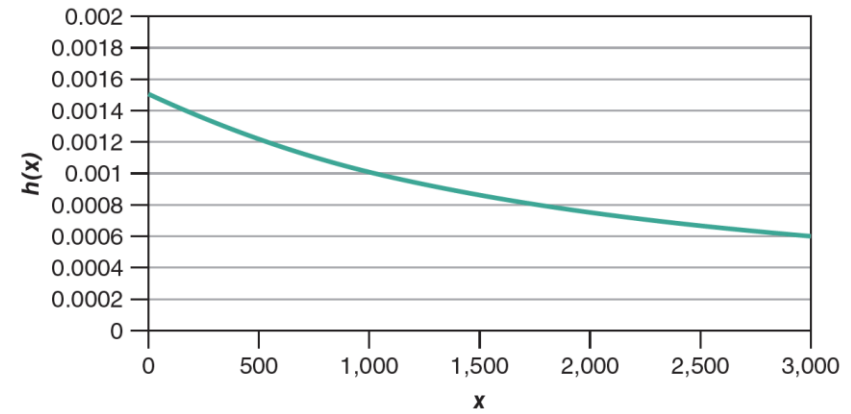
$$S(b) = \exp\left(-\int_0^b h(x) dx\right)$$

Уровень риска для Моделей

$$h_1(x) = \frac{0.01}{1 - 0.01x}, 0 < x < 100,$$



$$h_2(x) = \frac{3}{x + 2000}, x > 0,$$



$$h_3(x) \text{ не определен, } h_4(x) = 0.000001, x > 0.$$

Снова заметим, что для смешанного распределения уровень риска определяется только на части носителя. В точках дискретной функции массы вероятности уровень риска не определяется

Модель 5

- Модель иллюстрирует ситуацию, в которой функция плотности и уровень риска не определены в некоторой точке.
- Альтернативное распределение продолжительности жизни (по сравнению с Моделью 1).

$$S_5(x) = \begin{cases} 1 - 0.01x, & 0 \leq x < 50 \\ 1.5 - 0.02x, & 50 \leq x < 75 \end{cases}$$

- Функция кусочно-линейна, и производная в точке 50 не определена. Поэтому ни функция плотности, ни уровень риска не определены в точке 50. В отличие от смешанного распределения Модели 4, дискретной массы вероятности в этой точке тоже нет, поскольку вероятность 50 равна нулю (в силу непрерывности). В таких случаях можно доопределять произвольным образом функцию плотности или уровень риска.

Мода

- Еще одна характеристика случайной величины – это наиболее вероятное значение.
- **Определение 8.** Модой случайной величины называется наиболее вероятное значение.
Для дискретной переменной это значение с наибольшей вероятностью.
Для непрерывной величины это наибольшее значение функции плотности.
Если имеется несколько локальных максимумов функции плотности, они все считаются модами.
- **Модель 1:** функция плотности $f_1(x) = 0.01$, $0 < x < 100$, постоянна. Все значения от 0 до 100 могут быть модами, или, что то же самое, можно сказать, что моды нет
- **Модель 2:** Функция плотности $f_2(x) = \frac{3(2000)^3}{(x+2000)^4}$, $x > 0$, строго убывает, поэтому мода в т. 0

Моды в Моделях 3, 4 и 5

- **Модель 3:**

$$p_3(x) = \begin{cases} 0.5, x = 0 \\ 0.25, x = 1, \\ 0.12, x = 2, \\ 0.08, x = 3, \\ 0.05, x = 4 \end{cases}$$

- Наибольшую вероятность имеет значение 0.
- **Модель 4:** для смешанного распределения мода не определяется
- **Модель 5:**

$$f_5(x) = \begin{cases} 0.01, 0 \leq x < 50 \\ 0.02, 50 \leq x < 75 \end{cases}$$

- плотность имеет два участка постоянства, большее значение на интервале от 50 до 75, все эти значения являются модами.

Основные количественные характеристики распределений: моменты

- **Определение 9.** k -й момент случайной величины - это ожидаемое (среднее) значение k -й степени этой величины (если оно существует). Обозначается через $E[X^k]$ или μ'_k
- Первый момент – математическое ожидание, часто обозначается через μ .

! μ никак не связано с интенсивностью смертности $\mu(x)$

$$\mu'_k = E[X^k] = \begin{cases} \int_{-\infty}^{\infty} x^k f(x) dx, & \text{если сл. в. непрерывна} \\ \sum_j x_j^k p(x_j), & \text{если сл. в. дискретна} \end{cases}$$

- Все формулы приводятся для всюду непрерывных или всюду дискретных случайных величин. Для смешанных распределений соответствующие выражения получаются интегрированием по плотности распределения там, где переменная непрерывна, и суммированием по функции вероятности там, где переменная дискретна.

Первые два момента для Моделей 1-5

$$E[X_1] = \int_0^{100} x(0.01)dx = 50, E[X_1^2] = \int_0^{100} x^2(0.01)dx = 3333.33,$$

$$E[X_2] = \int_0^{\infty} x \frac{3(2000)^3}{(x+2000)^4} dx = 1000, E[X_2^2] = \int_0^{\infty} x^2 \frac{3(2000)^3}{(x+2000)^4} dx = 4000000,$$

$$E[X_3] = 0 \cdot 0.5 + 1 \cdot 0.25 + 2 \cdot 0.12 + 3 \cdot 0.08 + 4 \cdot 0.05 = 0.93,$$

$$E[X_3^2] = 0 \cdot 0.5 + 1 \cdot 0.25 + 4 \cdot 0.12 + 9 \cdot 0.08 + 16 \cdot 0.05 = 2.25$$

$$E[X_4] = 0 \cdot 0.7 + \int_0^{\infty} x \cdot 0.000003e^{-0.00001x} dx = 30000, E[X_4^2]$$

$$= 0^2 \cdot 0.7 + \int_0^{\infty} x^2 \cdot 0.000003e^{-0.00001x} dx = 6 \cdot 10^9,$$

$$E[X_5] = \int_0^{50} x \cdot 0.01 dx + \int_{50}^{75} x \cdot 0.02 dx = 43.75, E[X_5^2] = \int_0^{50} x^2 \cdot 0.01 dx + \int_{50}^{75} x^2 \cdot 0.02 dx = 2395.83$$

Основные количественные характеристики распределений: центральные моменты

- **Определение 10.** k -й центральный момент случайной величины равен математическому ожиданию k -й степени отклонения величины от среднего $E[(X - \mu)^k] \stackrel{\text{def}}{=} \mu_k$
- Второй центральный момент называется *дисперсией* σ^2 или $Var[X]$.
- Корень из дисперсии σ называется *стандартным отклонением*.
- Отношение стандартного отклонения к среднему значению называется *коэффициентом вариации*.
- Отношение третьего центрального момента к кубу стандартного отклонения называется *коэффициентом асимметрии*: $\gamma_1 = \frac{\mu_3}{\sigma^3}$.
- Отношение четвертого центрального момента к четвертой степени стандартного отклонения называется *эксцессом (куртозисом)* $\gamma_2 = \frac{\mu_4}{\sigma^4}$.

$$\mu_k = E[(X - \mu)^k] = \begin{cases} \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx, & \text{если сл. в. непрерывна} \\ \sum_j (x_j - \mu)^k p(x_j), & \text{если сл. в. дискретна} \end{cases}$$

- Интеграл берется только по тем x , для которых $f(x) > 0$.

Основные количественные характеристики распределений: центральные моменты

- Стандартное отклонение является мерой того, насколько вероятность распределена по значениям случайной величины. Оно имеет размерность случайной величины.
- Коэффициент вариации измеряет разброс относительно среднего.
- Коэффициент асимметрии является мерой асимметрии распределения. Для симметричных распределений он равен 0. Положительное значение показывает, что вероятности в правой части относятся к значениям, отстоящим дальше от среднего, чем в левой.
- Эксцесс является мерой «хвостовости» распределения относительно нормального распределения (для которого эксцесс равен 3). Эксцесс больше 3 показывает, что (при постоянном стандартном отклонении) относительно нормального распределения большую вероятность имеют точки, удаленные от среднего, чем точки, расположенные рядом со средним.
- Коэффициент вариации, коэффициент асимметрии и эксцесс являются безразмерными
- Связь между вторыми моментами:

$$\mu_2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx = E[X^2] - 2\mu E[X] + \mu^2 = \mu'_2 - \mu^2$$

Пример 3

Покажем, что функция плотности гамма-распределения имеет положительный коэффициент асимметрии.

гамма-функция $\Gamma(\alpha) = \int_0^{\infty} t^{-\alpha} e^{-t} dt$; неполная гамма-функция $\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{-\alpha} e^{-t} dt$

гамма-распределение:
$$F(x) = \Gamma\left(\alpha; \frac{x}{\theta}\right), \quad f(x) = \frac{\left(\frac{x}{\theta}\right)^{\alpha} e^{-x/\theta}}{x\Gamma(\alpha)}$$

k -й момент гамма-распределения: $E[X^k] = \frac{\theta^k \Gamma(\alpha+k)}{\Gamma(\alpha)} = \theta^k (\alpha + k - 1) \dots \alpha$

$$E[X] = \alpha\theta, E[X^2] = \alpha(\alpha + 1)\theta^2, E[X^3] = \alpha(\alpha + 1)(\alpha + 2)\theta^3, Var[X] = \alpha\theta^2, \mu_3 = 2\alpha\theta^3.$$

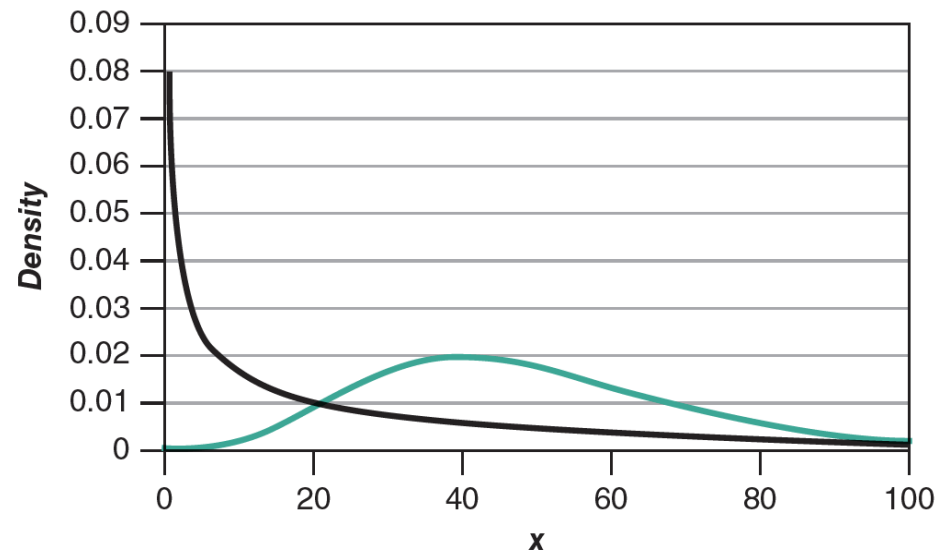
Коэффициент асимметрии равен $2\alpha^{-1/2}$. Так как $\alpha > 0$, то коэффициент асимметрии всегда положительный и убывает по α

Пример 3

Рассмотрим два гамма-распределения:

$$\alpha = 0.5, \theta = 100 (f(x)) \text{ и } \alpha = 5, \theta = 10 (g(x)).$$

$E[X] = 50$ в обоих случаях, но коэффициенты асимметрии равны 2.83 и 0.89, соответственно



Превышение ущерба

- **Определение 11.** Для заданного значения d такого, что $P\{X > d\} > 0$, превышение ущерба есть
$$Y^d = X - d \quad \text{для} \quad X > d$$

- Ожидаемое значение

$$e_X(d) = e(d) = E[Y^d] = E[X - d | X > d]$$

называется *функцией среднего превышения ущерба*. Другие названия: остаточное время жизни и полная ожидаемая продолжительность жизни.

- Y^d также называется усеченной слева и сдвинутой переменной.
- Усеченная слева – значит, что значения меньше d не рассматриваются. Сдвинутая, так как d вычитается из остальных значений.
- Когда X является переменной, отвечающей платежам, среднее превышение ущерба равно ожидаемому количеству заплаченного при условии, что платежи производятся после вычета d .
- Когда X отвечает возрасту смерти, среднее превышение ущерба является ожидаемым оставшимся временем до момента смерти при условии, что человек дожил до возраста d .

Превышение ущерба: моменты

- k -й момент величины превышения ущерба

$$e_X^k(d) = \begin{cases} \frac{\int_d^\infty (x-d)^k f(x) dx}{1-F(d)} & \text{для непрерывной переменной} \\ \frac{\sum_{x_j > d} (x_j - d)^k p(x_j)}{1-F(d)} & \text{для дискретной переменной} \end{cases}$$

- Интегрируя по частям первый момент, получаем более простое выражение:

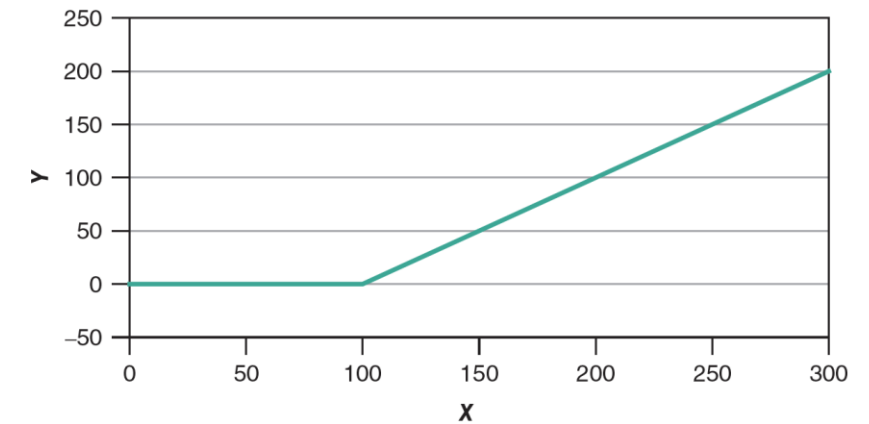
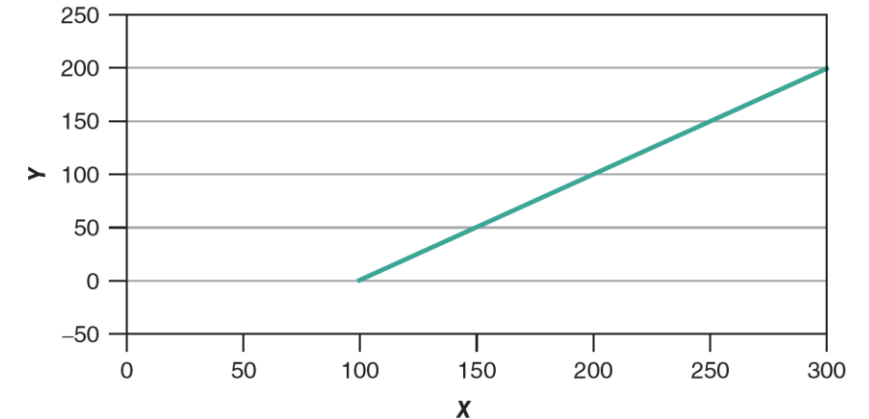
$$e_X(d) = \frac{\int_d^\infty (x-d)f(x)dx}{1-F(d)} = \frac{-(x-d)S(d)|_d^\infty + \int_d^\infty S(x)dx}{S(d)} = \frac{\int_d^\infty S(x)dx}{S(d)}$$

Цензурированные переменные

- **Определение 12.** Цензурированная слева и сдвинутая переменная

$$Y^L = (X - d)_+ = \begin{cases} 0, & X \leq d \\ X - d, & X > d \end{cases}$$

- Левостороннее цензурирование означает, что значения меньше d не игнорируются, а приравниваются 0.
- Стандартных обозначений для моментов таких переменных нет.
- Различие между переменными превышения и цензурированными слева переменными для финансовых потоков состоит в том, что первое рассчитывается для платежа, а второе для убытка.
- В расчете на платеж переменная существует только, если платежи были произведены. В расчете на ущерб переменная принимает нулевое значение всегда, когда ущерб не приводит к выплате.



Моменты цензурированной случайной величины

$$E[(X - d)_+^k] = \begin{cases} \int_d^{\infty} (x - d)^k f(x) dx & \text{для непрерывной переменной} \\ \sum_{x_j > d} (x_j - d)^k p(x_j) & \text{для дискретной переменной} \end{cases}$$

Не трудно видеть, что $E[(X - d)_+^k] = e^k(d)(1 - F(d))$

Пример 4

- Полис автомобильного страхования имеет следующие возможные варианты убытка (в скобках указана вероятность):

100 (0.4), 500 (0.2), 1,000 (0.2), 2,500 (0.1) и 10,000 (0.1).

Определить функцию массы вероятности и ожидаемые значения превышения ущерба и левостороннего цензурирования со сдвигом. Вычет равен 750.

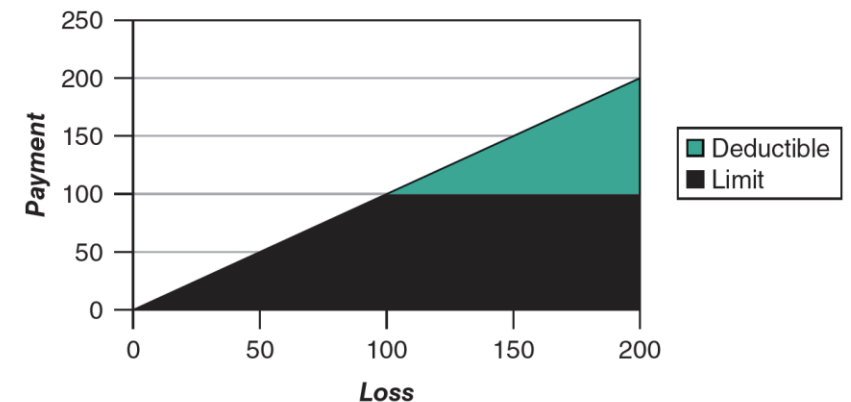
- Для превышения ущерба 750 вычитается из каждого возможного значения убытка.
- Получаем значения случайных величин 250, 1750 и 9250. Условная вероятность получает делением каждого из трех значений на вероятность превышения вычета, т.е. $1 - 0.6 = 0.4$. Получаем 0.5, 0.25 и 0.25. Математическое ожидание равно $250 \cdot 0.5 + 1750 \cdot 0.25 + 9250 \cdot 0.25 = 2875$.
- Для цензурированной слева и сдвинутой переменной вероятности значений меньше 750 присвоены нулевому значению. Остальные вероятности остаются без изменений, но значения, которым они приписаны, уменьшаются на величину вычета. Функция массы вероятности равна 0 (0.6), 250 (0.2), 1750 (0.1) и 9750 (0.1). Математическое ожидание $0 \cdot 0.6 + 250 \cdot 0.2 + 1750 \cdot 0.1 + 9250 \cdot 0.1 = 1150$.
- Отношение полученных математических ожиданий равно вероятности превышения вычета.
- Другой способ понять разницу между этими ожидаемыми величинами – рассмотреть 10 случаев с убытками, подчиняющимися указанному распределению.
- Только 4 из них приводят к выплатами, поэтому умножение на ожидаемый платеж получаем $4 \cdot 2875 = 11500$ или считая, что каждый платеж в среднем приводит к 1150 также получаем 11500.
- Таким образом, важно не то, какую переменную использовать, а то, чтобы она была использована подходящим образом.

Ограничение убытка

- **Определение 13.** Величина ограниченного убытка

$$Y = X \wedge u = \begin{cases} X, & X < u \\ u, & X \geq u \end{cases}$$

- $E[X \wedge u]$ - ограниченная ожидаемая величина.
- Переменная также называется цензурированной справа, поскольку значения, большие u , устанавливаются равными u .
- Этой ситуации в страховании соответствует лимит страхового покрытия, указанный в полисе (максимальная выплата)
- Очевидно, что
$$(X - d)_+ + (X \wedge u) = X$$
- Покупка полиса с лимитом ответственности d и другого с вычетом d эквивалентна покупке полного покрытия



Моменты ограниченной случайной величины

$$E[(X \wedge u)^k] = \begin{cases} \int_{-\infty}^u x^k f(x) dx + u^k [1 - F(u)] & \text{непрерывная сл. в.} \\ \sum_{x_j \leq u} x_j^k p(x_j) + u^k [1 - F(u)] & \text{дискретная сл. в.} \end{cases}$$

Другой вариант

$$\begin{aligned} E[(X \wedge u)^k] &= \int_{-\infty}^0 x^k f(x) dx + \int_0^u x^k f(x) dx + u^k [1 - F(u)] \\ &= x^k F(x) \Big|_{-\infty}^0 - \int_{-\infty}^0 kx^{k-1} F(x) dx - x^k S(x) \Big|_0^u + \int_0^u kx^{k-1} S(x) dx + u^k S(u) = - \int_{-\infty}^0 kx^{k-1} F(x) dx + \int_0^u kx^{k-1} S(x) dx \end{aligned}$$

При $k=1$ получаем

$$E[X \wedge u] = - \int_{-\infty}^0 F(x) dx + \int_0^u S(x) dx$$

Пример 4(продолжение)

- Вычислим функцию вероятности и ожидаемое значение переменной ограниченного ущерба с лимитом 750. И покажем, что сумма ожидаемых значений ограниченной и цензурированной слева сдвинутой переменных равна математическому ожиданию исходной случайной величины.
- Все значения большие или равные 750 заменяются на 750 и их вероятности суммируются.
- Таким образом, получаем 100 (0.4), 500 (0.2) и 750 (0.4).
- Математическое ожидание равно $100 \cdot 0.4 + 500 \cdot 0.2 + 750 \cdot 0.4 = 440$.
- Математическое ожидание исходной величины равно

$$100(0.4) + 500(0.2) + 1000(0.2) + 2500(0.1) + 10000(0.1) = 1590 = 440 + 1150$$

Основные количественные характеристики распределений: процентиля

- **Определение 14.** $100p$ процентиль случайной величины это любое значение π_p , такое, что

$$F(\pi_p -) \leq p \leq F(\pi_p).$$

- 50 процентиль, $\pi_{0.5}$, называется *медианой*.
- *Квантиль* принимает значения от 0 до 1, являясь по сути функцией, обратной к функции распределения.
- Процентиль принимает значения от 0 до 100, выражая вероятность в процентах (0.7 квантиль и 70 процентиль – это одно и то же).
- Если функция распределения имеет значение p для единственного значения x , то процентиль определяется однозначно.
- Если функция распределения имеет скачок от значения меньше p до значения больше p , то процентиль определяет в месте этого скачка.
- Процентиль не определяется однозначно, если функция распределения является константой со значением p для некоторого диапазона случайных величин. В этом случае любое значение из диапазона является процентилем.

Пример 5

- Определить 50 и 80 проценти для Моделей 1 и 3
- **Модель 1:** p процентиль находится из условия $p = F(\pi_p) = 0.01\pi_p$, откуда $\pi_p = 100p$.
- Поэтому требуемые проценти равны 50 и 80
- **Модель 3:** функция распределения равна 0.5 при $0 \leq x < 1$, таким образом, любое из значений от 0 до 1 является 50 процентилем.
- Для вычисления 80 проценти заметим, что в точке $x = 2$ функция распределения имеет скачок от 0.75 до 0.87, поэтому $\pi_{0.8} = 2$

