VOLGOGRAD STATE TECHNICAL UNIVERSITY

Машинное обучение и нейросетевые модели

Лекция 1. Введение в байесовское моделирование

Лектор: Кравченя Павел Дмитриевич

Волгоград 2025

VOLGOGRAD STATE TECHNICAL UNIVERSITY

План лекции

1. Основные понятия анализа, линейной алгебры и теории вероятностей.
2. Сущность байесовского моделирования.
3. Графические вероятностные модели. Байесовские сети.
4. Понятие d-разделимости.
5. Plate notation.
6. Пример использования байесовских сетей.
7. Тензоры в PyTorch, их атрибуты.
8. Вероятностные распределения в PyTorch, формы распределений.
9. Практические примеры работы с тензорами и распределениями в PyTorch.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей

Пусть задано *вероятностное пространство*: $(\Omega, \mathcal{F}, \mathbb{P})$, где Ω – множество *элементарных событий* $\omega \in \Omega$, $\Omega \subseteq \mathcal{F}$ – *сигма-алгебра подмножеств* множества Ω, $\mathbb{P}$ – *вероятностная мера*, заданная на элементах $\mathcal{F}$.

Понятие	Определение	Пример
Случайная величина	Величина, которая принимает значения *случайным образом* в зависимости от исхода эксперимента (*измеримая* функция от элементарного события): $X\colon\ \Omega \to \mathbb{R}, \qquad \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}\ \ \forall x \in \mathbb{R}.$	*Эксперимент*: бросок игрального кубика. Пусть X – *случайная величина*, равная 14, если выпало *чётное* число. $\Omega = \{1, 2, \dots 6\}, \qquad X(\omega) = \begin{cases} 0, \text{если } \omega = 1, 3, 5; \\ 14, \text{если } \omega = 2, 4, 6. \end{cases}$
Реализация случайной величины	*Определённое* значение случайной величины, которое наблюдается в конкретном эксперименте: $x = X(\omega), \qquad \omega \in \Omega.$	После *броска кубика* выпала «двойка» *Элементарный исход*: $\omega = 2$. *Реализация случайной величины*: $x = X(2) = 14.$
Распределение случайной величины	Закон, описывающий, насколько вероятно появление конкретной реализации СВ (на некотором *борелевском* множестве): $\mu_X(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})\ \ \forall B \in \mathfrak{B}(\mathbb{R}).$	*Борелевское* множество $B = \{14\}$. $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in \{14\}\} = \{2, 4, 6\};$ $\mu_X(B) = \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}(\{2, 4, 6\}) = 0{,}5.$ *Аналогично* для $B = \{0\}$.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей. Описание случайных величин

В зависимости от *области значений*, которые может принимать случайная величина, случайные величины разделяются на *дискретные* и *непрерывные* (а также *смешанные*).

Понятие	Дискретная случайная величина	Непрерывная случайная величина
Функция вероятности (PMF, Probability Mass Function).	$\mathrm{P}(X = x_i) = p(x_i) = \mathrm{P}_i$. Может быть задана *таблицей*.	$\mathrm{P}(X = x) = 0$. *Не используется* для описания!
Функция распределения (CDF, Cumulative Distribution Function).	$F(x) = F_X(x) = \mathrm{P}(X \le x) = \sum_{i:x_i \le x} p(x_i)$. $F(x)$ неубывающая, $0 \le F(x) \le 1$.	$F(x) = F_X(x) = \mathrm{P}(X \le x) = \int_{-\infty}^{x} p(t)dt$. $F(x)$ неубывающая, $0 \le F(x) \le 1$.
Плотность функции распределения (PDF, Probability Density Function).	*Не определена* в силу дискретности СВ.	$p(x) = \lim_{\Delta x \to +0} \frac{\mathrm{P}(x < X \le \mathrm{x} + \Delta x)}{\Delta x}$, $\mathrm{P}(a < X \le b) = F(b) - F(a)$, $p(x) \ge 0$.
Условие нормировки	$\sum_i p(x_i) = 1$.	$\int_{-\infty}^{+\infty} p(x)dx = 1$.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей. Математическое ожидание и дисперсия

В байесовском анализе особую роль играют *математическое ожидание* и *дисперсия функции от случайной величины*, дающие представление о её *центре* и *разбросе*.

Понятие	Дискретная случайная величина	Непрерывная случайная величина
Математическое ожидание	$\mathbb{E}_{X\sim F_X}[f(X)] = \mathbb{E}[f(X)] = \sum_i f(x_i)\cdot \mathrm{P}(X = x_i)\,.$	$\mathbb{E}_{X\sim p(x)}[f(X)] = \mathbb{E}[f(X)] = \int_{-\infty}^{+\infty} f(x)\cdot p(x)dx\,.$
Дисперсия	$D_{X\sim F_X}[f(X)] = D[f(X)] = \mathbb{E}[(f(X) - \mathbb{E}[f(X)])^2],$ $D[f(X)] = \sum_i (f(x_i) - \mathbb{E}[f(X)])^2 \cdot \mathrm{P}(X = x_i)\,,$ $D[f(X)] = \mathbb{E}[f(X)^2] - (\mathbb{E}[f(X)])^2 =$ $= \sum_i f(x_i)^2 \cdot \mathrm{P}(X = x_i) - (\mathbb{E}[f(X)])^2.$	$D_{X\sim p(x)}[f(X)] = D[f(X)] = \mathbb{E}[(f(X) - \mathbb{E}[f(X)])^2],$ $D[f(X)] = \int_{-\infty}^{+\infty} (f(x) - \mathbb{E}[f(X)])^2 \cdot p(x)dx\,,$ $D[f(X)] = \mathbb{E}[f(X)^2] - (\mathbb{E}[f(X)])^2 =$ $= \int_{-\infty}^{+\infty} f(x)^2 \cdot p(x)dx - (\mathbb{E}[f(X)])^2.$
Среднеквадратичное отклонение	$\sigma_{X\sim F_X}[f(X)] = \sigma[f(X)] = \sqrt{D[f(X)]}.$	$\sigma_{X\sim p(x)}[f(X)] = \sigma[f(X)] = \sqrt{D[f(X)]}.$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей. Совместное распределение

- *Совместное распределение случайных величин* описывает вероятностное поведение *нескольких величин одновременно*. Оно описывается:

  ✓ Для дискретных распределений: *совместной функции вероятности* (Joint Probability Mass Function):

$$\mathrm{P}(X, Y) = \mathrm{P}(X = x, Y = y).$$

  ✓ Для непрерывных распределений: *совместной функцией плотности вероятности* (Joint Probability Density Function):

$$p(x, y) = \lim_{\substack{\Delta x \to +0 \\ \Delta y \to +0}} \frac{P(x < X \leq \mathrm{x} + \Delta x, y < Y \leq \mathrm{y} + \Delta y)}{\Delta x \Delta y}.$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Некоторые понятия анализа и алгебры

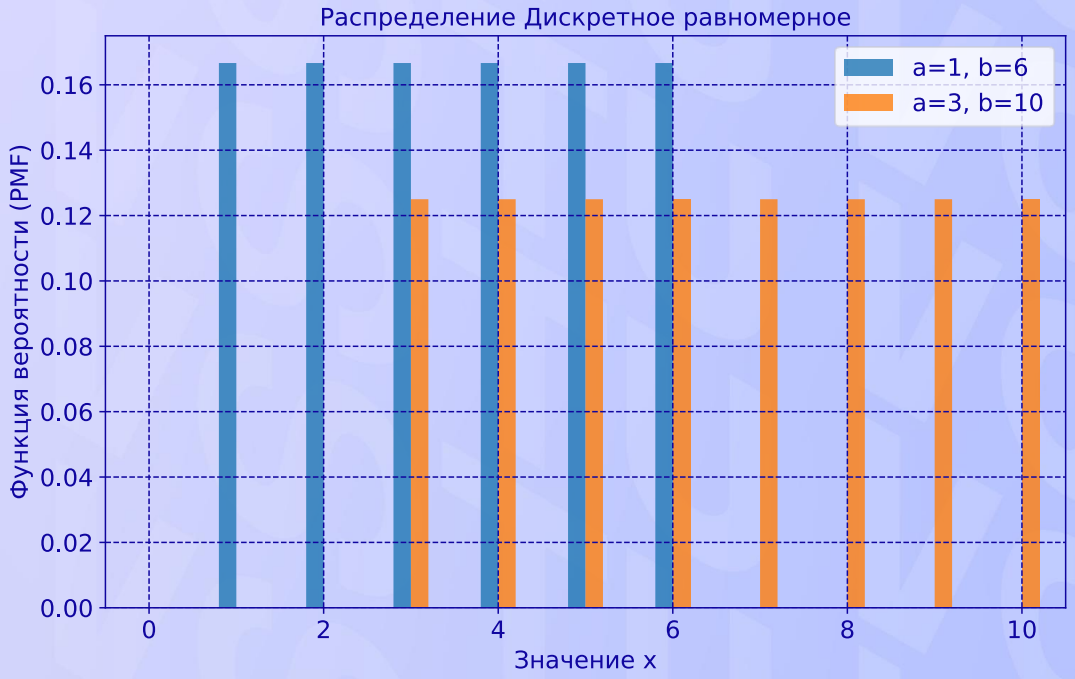
Для анализа свойств распределений часто используются следующие математические понятия.

Понятие	Обозначение	Определение
Стандартный $(k-1)$*-мерный симплекс*	Δ^{k-1}	Множество точек в k-мерном пространстве, координаты которых удовлетворяют *следующим условиям*: $\Delta^{k-1} = \left\{ (x_1, x_2, \dots, x_k) \in \mathbb{R}^k : \forall i \in [1..k]\ x_i \geq 0 \text{ и } \sum_{i=1}^{k} x_i = 1 \right\}.$
Носитель функции	$\operatorname{supp}(u)$	Замыкание множества X, на котором *вещественнозначная* функция $u: X \to \mathbb{R}$ *не обращается в нуль*: $\operatorname{supp}(u) = \overline{\{x : u(x) \neq 0\}}.$ Это множество значений, которые случайная величина может принимать с *ненулевой вероятностью* (или CDF).
Линейная оболочка множества векторов	$\operatorname{span}(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\})$	Множество *всех возможных линейных комбинаций* векторов: $\operatorname{span}(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}) = \left\{ \sum_{i=1}^{k} c_i \mathbf{x}_i : \forall i \in [1..k]\ c_i \in \mathbb{R} \right\}.$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Равномерное распределение

Характеристика	Значение
Обозначение	$\mathcal{D}(x; a, b)$
Параметры	$a \in \mathbb{N}$ – левая граница распределения; $b \in \mathbb{N}$ – правая граница, $b > a$.
Носитель (supp)	$[a, b] \cap \mathbb{N}$
Функция вероятности	$\mathrm{P}(X = x) = \frac{1}{b - a}$
Матожидание	$\mathbb{E}_{X \sim \mathcal{D}(x;a,b)}[X] = \frac{a + b}{2}$
Дисперсия	$D_{X \sim \mathcal{D}(x;a,b)}[X] = \frac{(b - a + 1)^2 - 1}{12}$

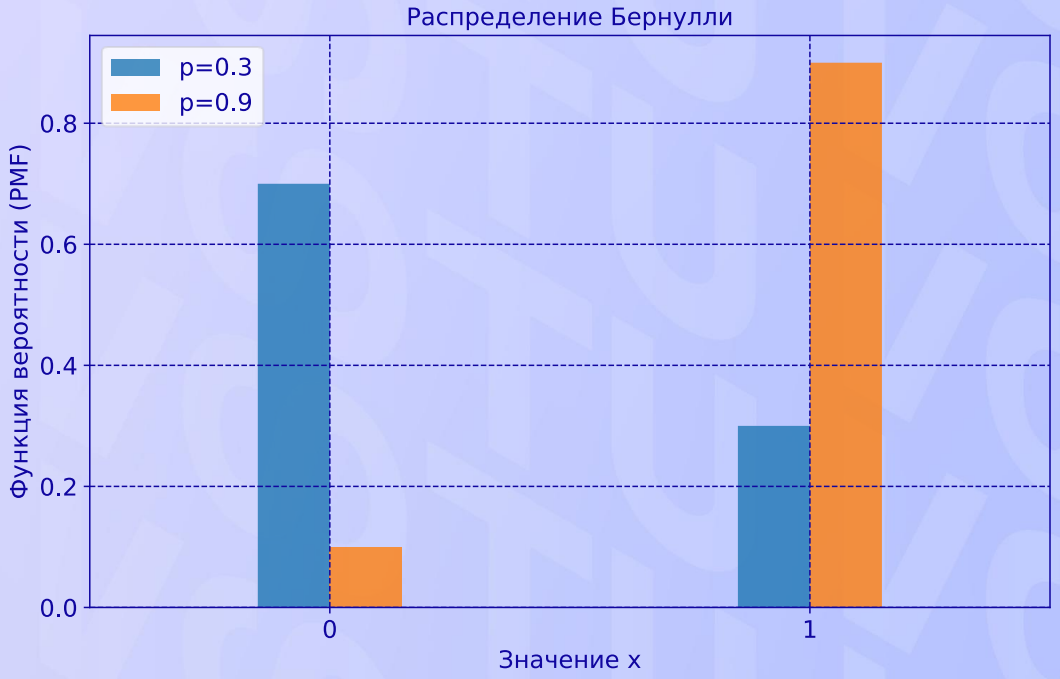


В дискретном равномерном распределении случайная величина может принимать *конечное число целочисленных значений с одинаковой вероятностью*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Распределение Бернулли

Характеристика	Значение
Обозначение	$\mathcal{B}(x;p)$
Параметры	$p \in (0,1)$
Носитель (supp)	$\{0,1\}$
Функция вероятности	$\mathrm{P}(X=x)=\begin{cases} p, & \text{если } x=1; \\ 1-p, & \text{если } x=0. \end{cases}$
Матожидание	$\mathbb{E}_{X\sim\mathcal{B}(x;p)}[X]=p$
Дисперсия	$D_{X\sim\mathcal{B}(x;p)}[X]=p(1-p)$

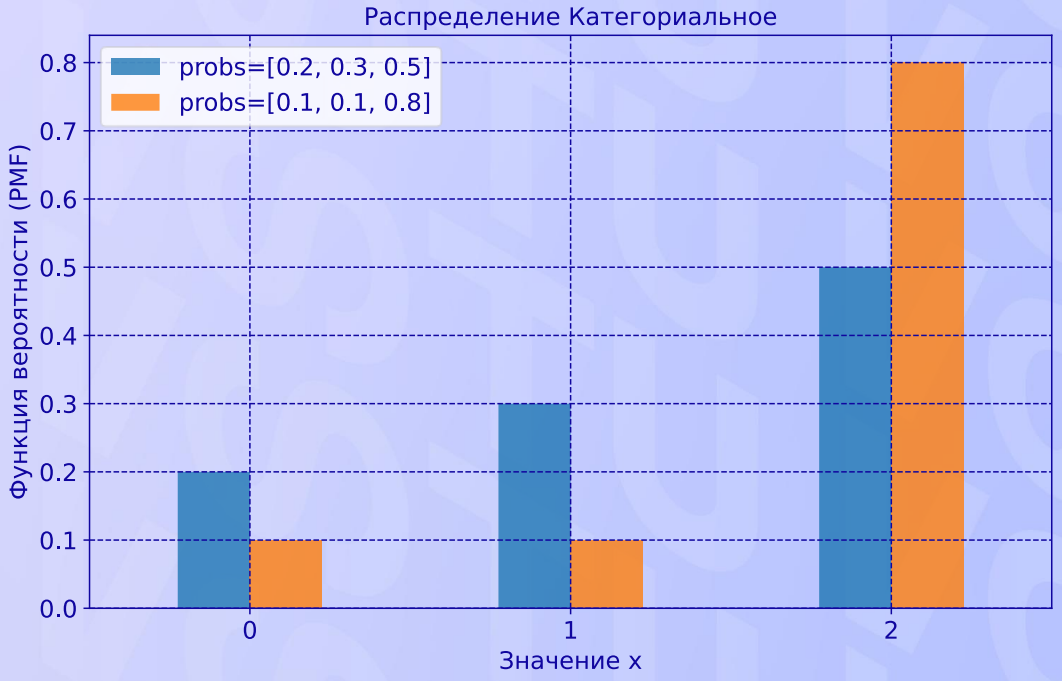


Распределение Бернулли описывает вероятности значений *бинарной случайной величины*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Категориальное распределение

Характеристика	Значение
Обозначение	$\mathcal{C}at(x; k, \boldsymbol{p})$
Параметры	$\boldsymbol{p} \in \Delta^{k-1}$
Носитель (supp)	$\{0, 1, \ldots, k-1\}$
Функция вероятности	$\mathrm{P}(X = x) = \begin{cases} p_0, & \text{если } x = 0; \\ \ldots & \\ p_{k-1}, & \text{если } x = k-1. \end{cases}$
Матожидание	$\mathbb{E}_{X \sim \mathcal{C}at(x;\boldsymbol{p})}[X] = \boldsymbol{p}$
Дисперсия	$D_{X \sim \mathcal{C}at(x;\boldsymbol{p})}[X] = \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}$

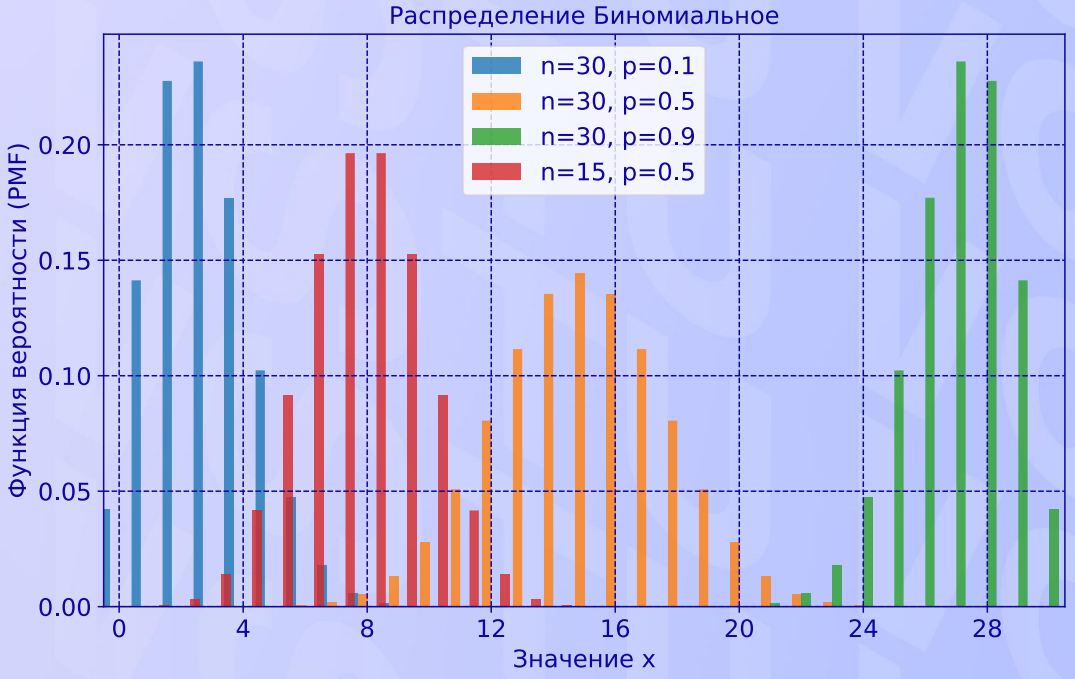


Категориальное распределение описывает вероятности значений *случайной величины*, равные одной из *нескольких возможных категорий*. Является *обобщением распределения Бернулли* для дискретной случайной величины с числом *исходов больше двух*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Биномиальное распределение

Характеристика	Значение
Обозначение	$\mathcal{B}in(x; n, p)$
Параметры	$n \in \mathbb{N}$ – количество испытаний, $p \in (0, 1)$ – вероятность успеха.
Носитель (supp)	$\{0, 1, \dots, n\}$
Функция вероятности	$\mathrm{P}(X = x) = C_n^x p^x (1-p)^{n-x}$
Матожидание	$\mathbb{E}_{X \sim \mathcal{B}in(x;n,p)}[X] = np$
Дисперсия	$D_{X \sim \mathcal{B}in(x;n,p)}[X] = np(1-p)$

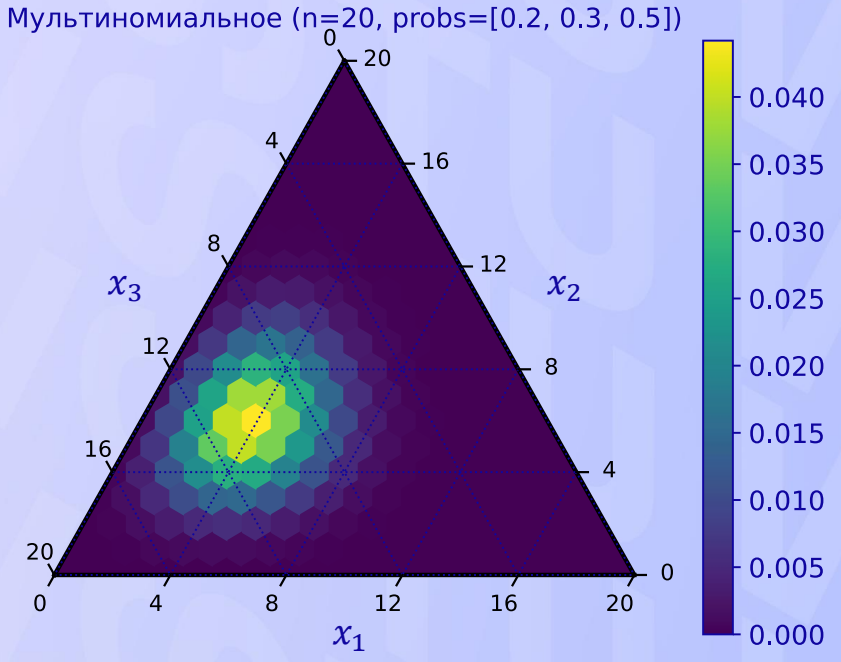


Биномиальное распределение описывает *распределение числа успехов* в серии из n экспериментов, каждый из которых завершается успехом с *вероятностью* p.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Мультиномиальное распределение

Характеристика	Значение
Обозначение	$\mathcal{Mult}(x; k, n, \boldsymbol{p})$
Параметры	$n \in \mathbb{N}$ – количество испытаний, $k \in \mathbb{N}$ – число исходов одного испыт. $\boldsymbol{p} \in \Delta^{k-1}$ – вероятности исходов.
Носитель (supp)	$n\Delta^{n-1} \cap \mathbb{Z}^k$
Функция вероятности	$\mathrm{P}(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdot \ldots \cdot x_k!} \cdot p_0^{x_1} \cdot \ldots \cdot p_0^{x_1}$
Матожидание	$\mathbb{E}_{X \sim \mathcal{Mult}(x;k,n,\boldsymbol{p})}[X] = n\boldsymbol{p}$
Дисперсия	$D_{X \sim \mathcal{Mult}(x;k,n,\boldsymbol{p})}[X_i] = np_i(1-p_i)$

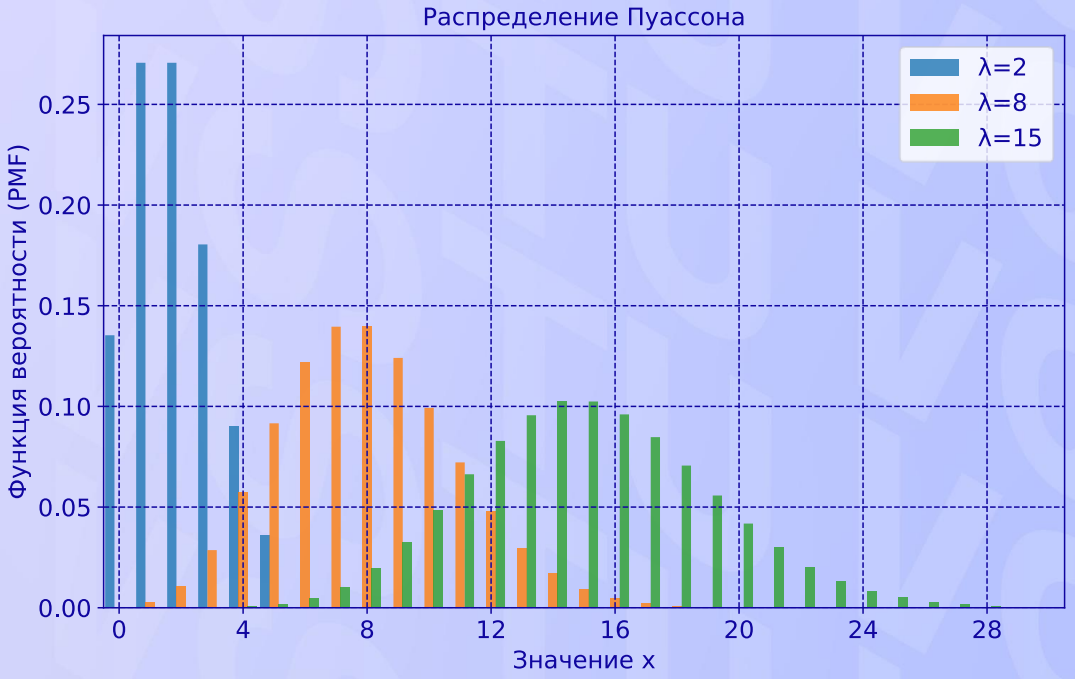


Мультиномиальное распределение описывает вероятность *получить определённый набор исходов* в серии из n экспериментов, каждое из которых имеет k исходов с *вероятностями* $p_1, p_2, \ldots, p_k$. Является *обобщением биномиального распределения*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Распределение Пуассона

Характеристика	Значение
Обозначение	$\mathcal{P}ois(x;\lambda)$
Параметры	$\lambda \in \mathbb{R}_+ -$ среднее число событий.
Носитель (supp)	$\mathbb{Z}_+$
Функция вероятности	$\mathrm{P}(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}$
Матожидание	$\mathbb{E}_{X\sim\mathcal{P}ois(x;\lambda)}[X]=\lambda$
Дисперсия	$D_{X\sim\mathcal{P}ois(x;\lambda)}[X]=\lambda$

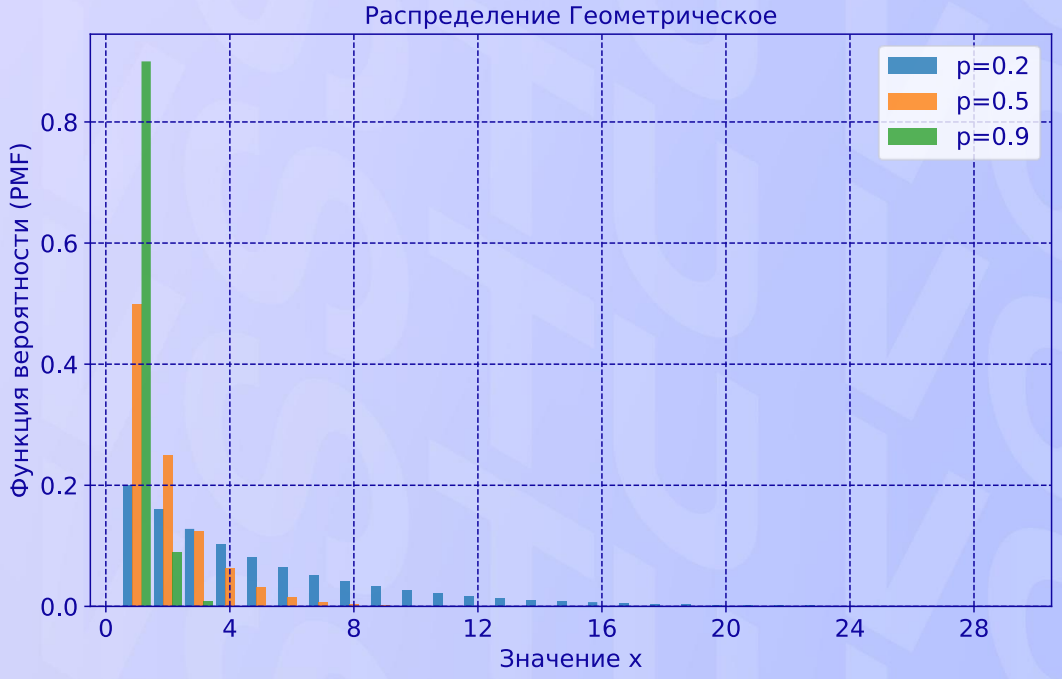


Распределение Пуассона описывает вероятность количества *редких событий*, происходящих в течение *фиксированного* интервала времени / области пространства, при условии, что эти события происходят *независимо друг от друга* и с *постоянной средней интенсивностью*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Дискретные распределения. Геометрическое распределение

Характеристика	Значение
Обозначение	$Geom(x;p)$
Параметры	$p \in (0,1]$ – вероятность успеха.
Носитель (supp)	$\mathbb{N}$
Функция вероятности	$\mathrm{P}(X=x)=(1-p)^{x-1}p$
Матожидание	$\mathbb{E}_{X\sim Geom(x;p)}[X]=\frac{1}{p}$
Дисперсия	$D_{X\sim Geom(x;p)}[X]=\frac{1-p}{p^2}$

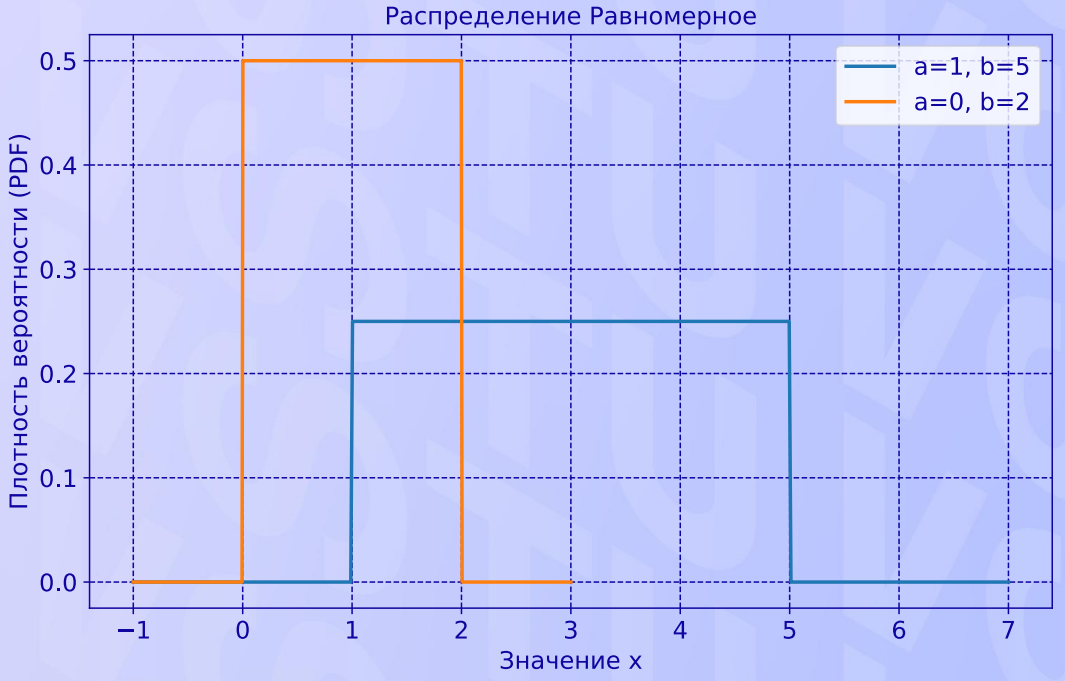


Геометрическое распределение описывает количество *неудач* до *первого успеха* в серии *испытаний Бернулли*, проводимых с *одинаковой вероятностью успеха* в каждом.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Равномерное распределение

Характеристика	Значение
Обозначение	$\mathcal{U}(x; a, b)$
Параметры	$a \in \mathbb{R}$ – левая граница распределения; $b \in \mathbb{R}$ – правая граница, $b > a$.
Носитель (supp)	$[a, b]$
Плотность вероятности	$p(x) = \frac{1}{b-a} \cdot \mathbb{I}[x \in [a, b]]$
Матожидание	$\mathbb{E}_{X \sim \mathcal{U}(x;a,b)}[X] = \frac{a+b}{2}$
Дисперсия	$D_{X \sim \mathcal{U}(x;a,b)}[X] = \frac{(b-a)^2}{12}$

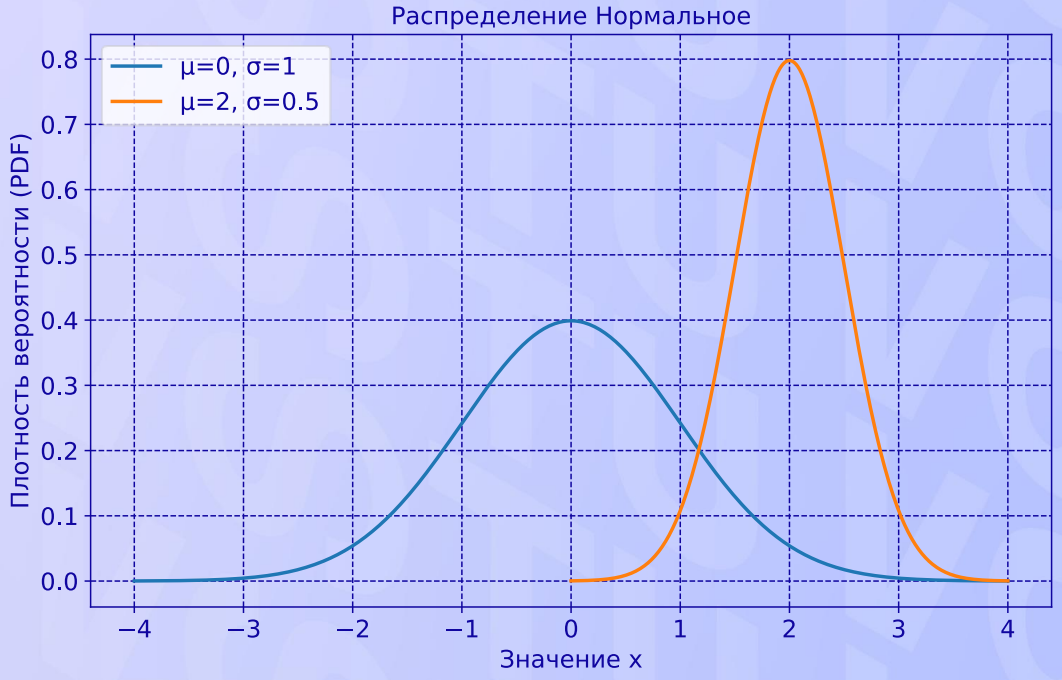


В непрерывном равномерном распределении случайная величина может принимать *любое значение из некоторого отрезка с одинаковой вероятностью*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Одномерное нормальное распределение

Характеристика	Значение
Обозначение	$\mathcal{N}(x;\mu,\sigma^2)$
Параметры	$\mu \in \mathbb{R}$ – матожидание; $\sigma^2 \in \mathbb{R}_+$ – дисперсия.
Носитель (supp)	$\mathbb{R}$
Плотность вероятности	$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
Матожидание	$\mathbb{E}_{X\sim\mathcal{N}(x;\mu,\sigma^2)}[X] = \mu$
Дисперсия	$D_{X\sim\mathcal{N}(x;\mu,\sigma^2)}[X] = \sigma^2$

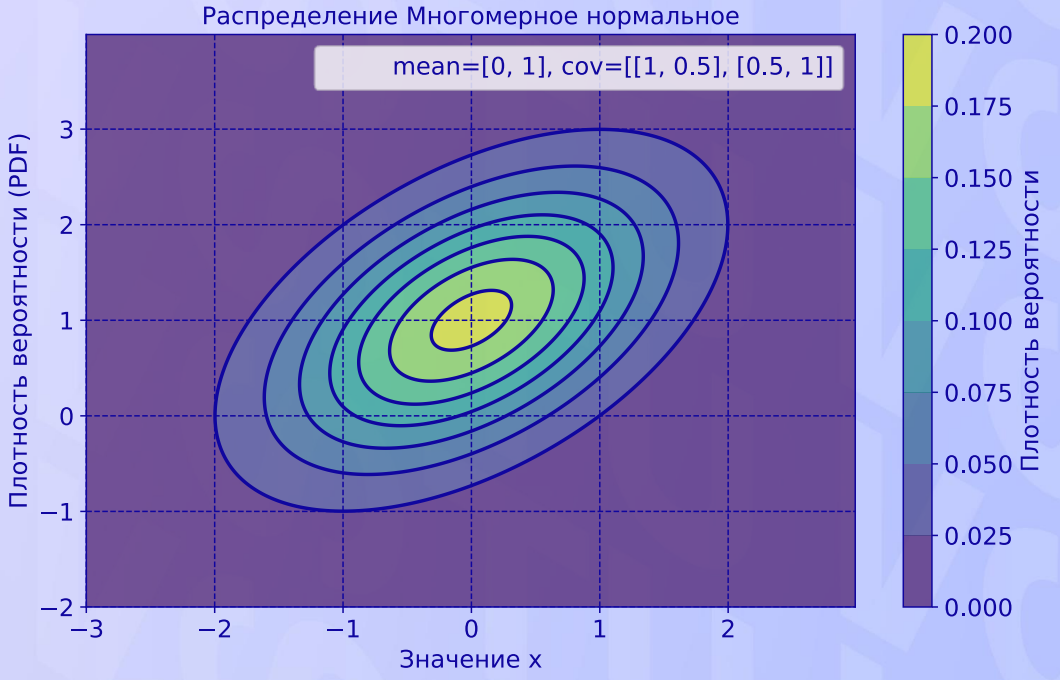


Нормальное распределение играет важную роль в связи с ЦПТ. Она утверждает, что сумма большого числа независимых случайных величин, *независимо от их исходного распределения*, при определенных условиях приближается к нормальному распределению.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Многомерное нормальное распределение

Характеристика	Значение
Обозначение	$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$
Параметры	$\boldsymbol{\mu} \in \mathbb{R}^n$ — вектор матожиданий; $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ — матрица ковариаций.
Носитель (supp)	$\boldsymbol{\mu} + \mathrm{span}(\boldsymbol{\Sigma}) \subseteq \mathbb{R}^n$
Плотность вероятности	$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \lvert\boldsymbol{\Sigma}\rvert^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$
Матожидание	$\mathbb{E}_{X \sim \mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})}[X] = \boldsymbol{\mu}$
Дисперсия	$D_{X \sim \mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})}[X] = \boldsymbol{\Sigma}$

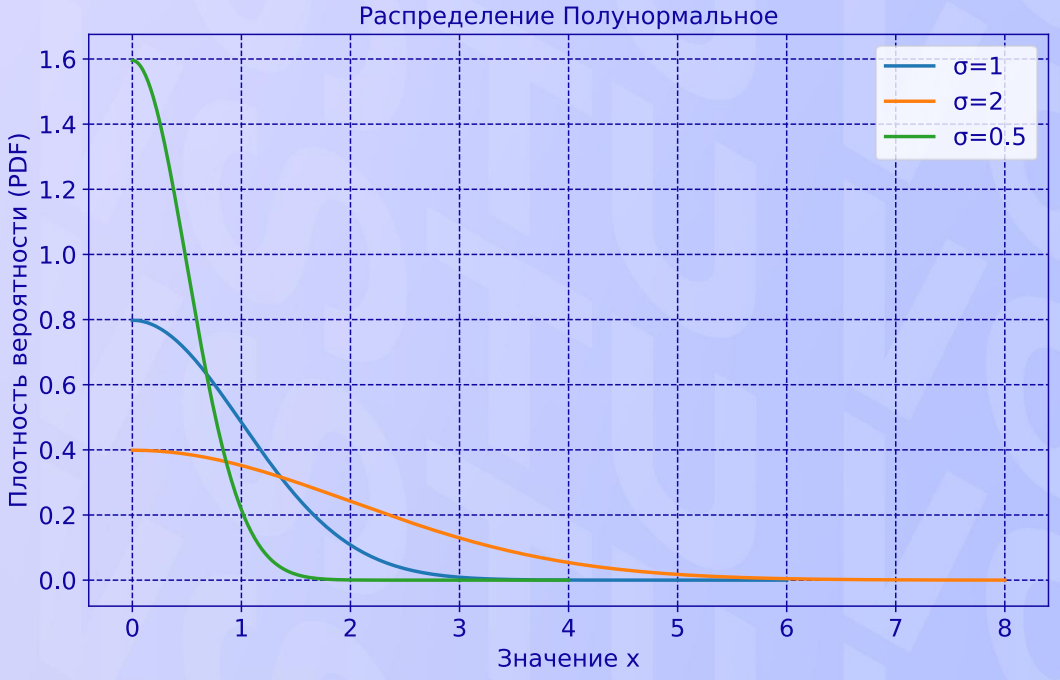


Многомерное нормальное распределение представляет собой *обобщение одномерного нормального распределения* для *нескольких* случайных величин, которые могут быть *зависимы друг от друга*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Полунормальное распределение

Характеристика	Значение
Обозначение	$\mathcal{HN}(x;\sigma)$
Параметры	$\sigma > 0$ – параметр масштаба.
Носитель (supp)	$\mathbb{R}_+ \cup \{0\}$
Плотность вероятности	$p(x) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \cdot \exp\left[-\frac{x^2}{2\sigma^2}\right] \cdot \mathbb{I}[x \geq 0]$
Матожидание	$\mathbb{E}_{X\sim\mathcal{HN}(x;\sigma)}[X] = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$
Дисперсия	$D_{X\sim\mathcal{HN}(x;\sigma)}[X] = \sigma^2\left(1-\frac{2}{\pi}\right)$

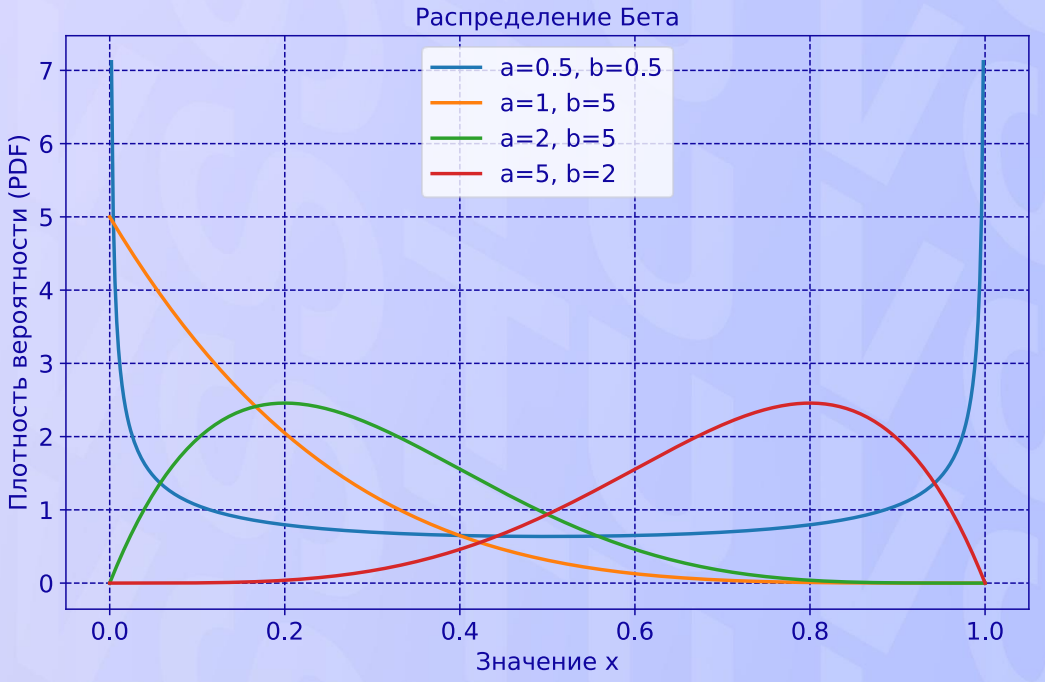


Полунормальное распределение – это распределение *абсолютного значения нормально распределенной величины с нулевым средним*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Бета-распределение

Характеристика	Значение
Обозначение	$Beta(x;\alpha,\beta)$
Параметры	$\alpha > 0;$ $\beta > 0.$
Носитель (supp)	$[0,1]$
Плотность вероятности	$p(x) = \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1}$
Матожидание	$\mathbb{E}_{X\sim Beta(x;\alpha,\beta)}[X] = \frac{\alpha}{\alpha+\beta}$
Дисперсия	$D_{X\sim Beta(x;\alpha,\beta)}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

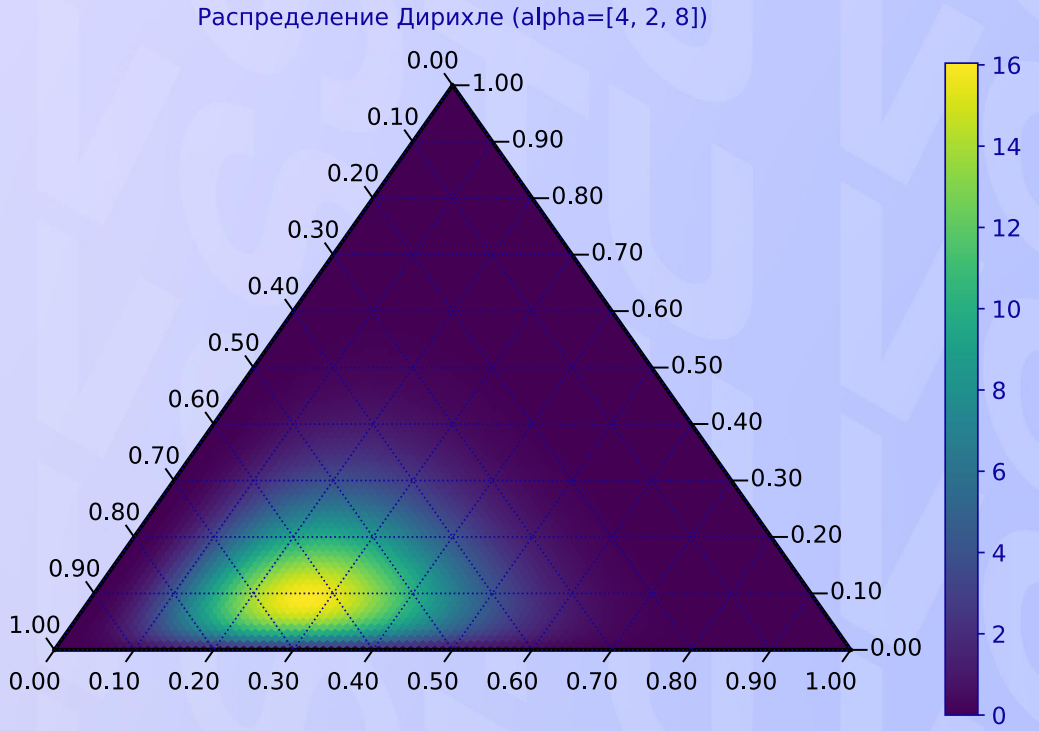


Бета-распределение часто используется для *моделирования величин, принимающих значения от нуля до единицы* (доли, пропорции, проценты, ...).

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Распределение Дирихле

Характеристика	Значение
Обозначение	$\mathcal{D}ir(\mathbf{x};\boldsymbol{\alpha})$
Параметры	$\boldsymbol{\alpha}$ – параметр концентрации; $\alpha_i > 0 \quad \forall i \in [1..n]$.
Носитель (supp)	Δ^{n-1}
Плотность вероятности	$p(x) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{n}\Gamma(\alpha_i)}\prod_{i=1}^{n} x_i^{\alpha_i-1}, \alpha_0 = \sum_{i=1}^{n}\alpha_i\,.$
Матожидание	$\mathbb{E}_{X\sim\mathcal{D}ir(\mathbf{x};\boldsymbol{\alpha})}[X_i] = \frac{\alpha_i}{\alpha_0}$
Дисперсия	$D_{X\sim\mathcal{D}ir(\mathbf{x};\boldsymbol{\alpha})}[X] = \frac{\mathbb{E}[X_i](1-\mathbb{E}[X_i])}{\alpha_0+1}$

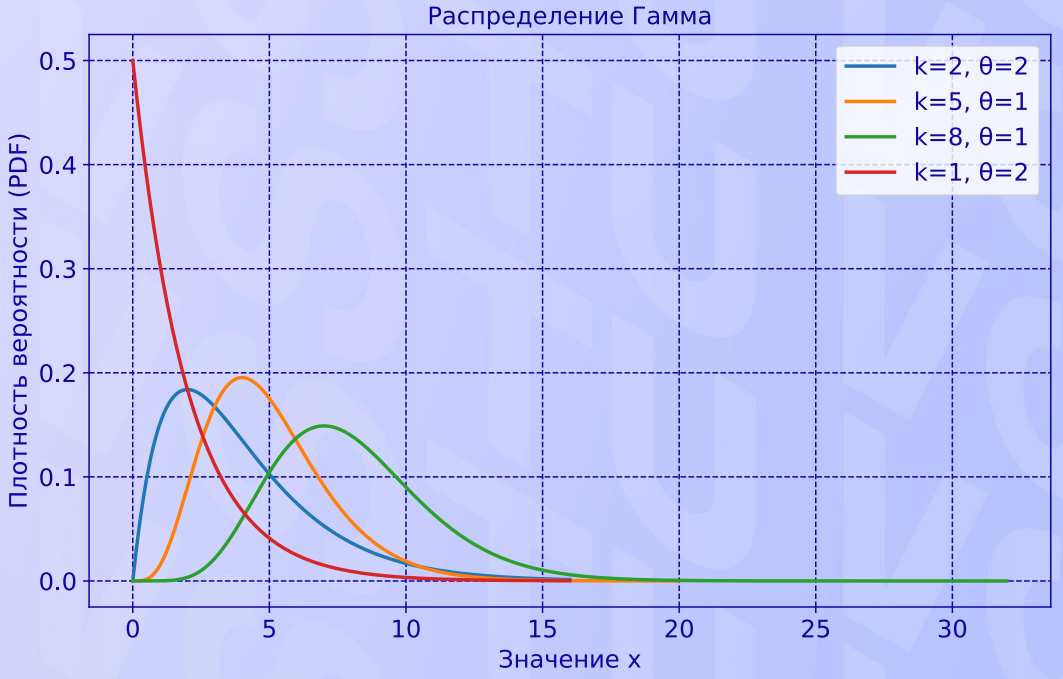


Распределение Дирихле часто используется для *моделирования вероятностей взаимоисключающих категорий*. Является *обобщением бета-распределения*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Гамма-распределение

Характеристика	Значение
Обозначение	$\mathcal{G}(x; k, \theta)$
Параметры	$k > 0$ – параметр формы; $\theta > 0$ – параметр масштаба.
Носитель (supp)	$\mathbb{R}_+$
Плотность вероятности	$p(x) = \frac{1}{\Gamma(k)\theta^k} \cdot x^{k-1} e^{-\frac{x}{\theta}}$
Матожидание	$\mathbb{E}_{X \sim \mathcal{G}(x;k,\theta)}[X] = k\theta$
Дисперсия	$D_{X \sim \mathcal{G}(x;k,\theta)}[X] = k\theta^2$

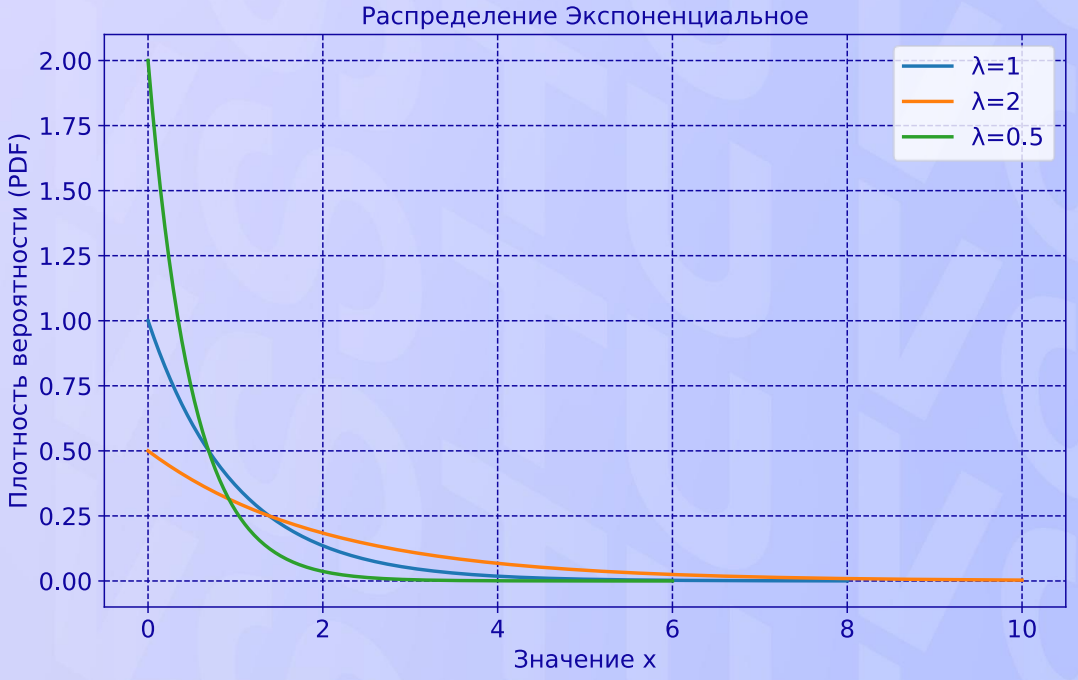


Гамма-распределение часто используется для *моделирования времени до наступления какого-либо события*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Экспоненциальное распределение

Характеристика	Значение
Обозначение	$\mathcal{E}xp(x;\lambda)$
Параметры	$\lambda > 0$ – интенсивность
Носитель (supp)	$\mathbb{R}_+$
Плотность вероятности	$p(x) = \lambda \cdot e^{-\lambda x}$
Матожидание	$\mathbb{E}_{X\sim\mathcal{E}xp(x;\lambda)}[X] = \frac{1}{\lambda}$
Дисперсия	$D_{X\sim\mathcal{E}xp(x;\lambda)}[X] = \frac{1}{\lambda^2}$

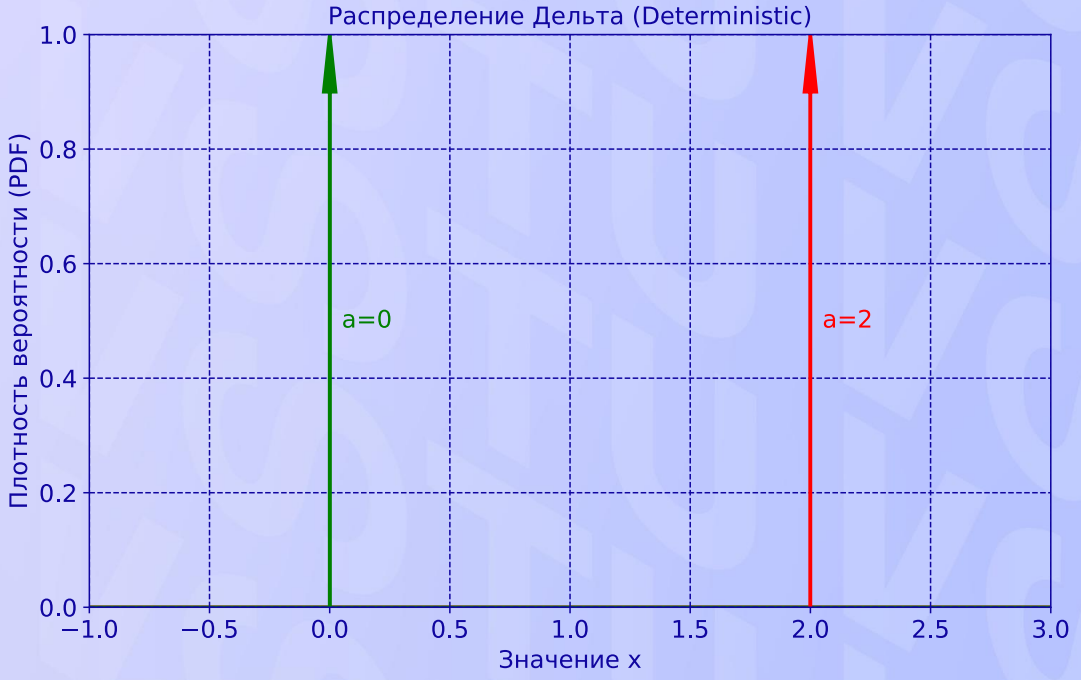


Гамма-распределение часто используется для *моделирования времени между событиями в процессе Пуассона* (в процессе, в котором события происходят *непрерывно и независимо* с *постоянной средней интенсивностью*).

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Непрерывные распределения. Детерминированное распределение

Характеристика	Значение
Обозначение	$\delta(x; a)$
Параметры	$a \in \mathbb{R}$ – детерминированное значение
Носитель (supp)	$\{a\}$
Плотность вероятности	$p(x) = \delta(x - a)$, $\delta(x) = \begin{cases} 0, & \text{если } x \neq a; \\ \infty, & \text{если } x = a. \end{cases}$
Матожидание	$\mathbb{E}_{X \sim \delta(x;a)}[X] = a$
Дисперсия	$D_{X \sim \delta(x;a)}[X] = 0$



Детерминированное распределение описывают ситуацию, когда *случайная величина по факту не является случайной*, а всегда принимает *одно и то же фиксированное значение*.

Условной плотностью вероятности $p(x|y)$ (conditional distribution) называют вероятность случайной величины X принять значение в некоторой малой окрестности x при условии того, что случайная величина Y приняла значение y:

$$p(x|y) = \frac{p(x,y)}{p(y)} \qquad conditional = \frac{joint}{marginal}$$

или:

$$p(x,y) = p(x|y)p(y)$$

Условное распределение показывает, как распределена X, если известна Y.

Если $p(x|y) = p(x)$, то случайные величины X и Y называются **независимыми**. В этом случае:

$$p(x,y) = p(x)p(y)$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей. Правило произведения

Правило произведения является обобщением связи между совместным и условными распределениями:

$$p(x_1, x_2) = p(x_2|x_1) \cdot p(x_1)$$

$$p(x_1, x_2, x_3) = p(x_3|x_1, x_2) \cdot p(x_2|x_1) \cdot p(x_1)$$

Или, в общем случае:

$$p(x_1, x_2, \dots x_n) = p(x_n|x_1, x_2, \dots x_{n-1}) \cdot p(x_{n-1}|x_1, x_2, \dots x_{n-2}) \dots p(x_2|x_1) \cdot p(x_1)$$

$$p(x_1, x_2, \dots x_n) = p(x_1) \prod_{i=2}^{n} p(x_i|x_{i-1}, \dots, x_1)$$

Последовательность записи переменных не играет роли:

$$p(x_1, x_2, \dots x_n) = p(x_1|x_2, x_3, \dots x_n) \cdot p(x_2|x_3, x_2, \dots x_n) \dots p(x_{n-1}|x_n) \cdot p(x_n)$$

Обращение условного распределения:

$$p(x,y) = p(x|y) \cdot p(y) = p(y|x) \cdot p(x)$$

Отсюда:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Проинтегрируем обе части выражения по y:

$$\int p(y|x)dy = \frac{1}{p(x)} \int p(x|y)p(y)dy$$

Применяя правило нормировки, получаем **правило суммы**:

$$p(x) = \int p(x|y)p(y)dy = \int p(x,y)dy$$

Выражение для **маргинализации** случайной величины можно получить из правила *суммирования* и *определения матожидания* функции:

$$p(x) = \int p(x, y)dy = \int p(x|y)p(y)dy = \mathbb{E}_{y\sim p(y)}[p(x|y)]$$

Аналогично, если задано совместное распределение n случайных величин $p(x_1, x_2, \dots x_n)$, то **маргинальное** (безусловное) распределение k из них ($k < n$):

$$p(x_1, x_2, \dots x_k) = \int p(x_1, x_2, \dots x_n)dx_{k+1}dx_{k+2} \cdots dx_n$$

Аналогичное выражение для дискретных случайных величин:

$$p(x_1, x_2, \dots x_k) = \sum_{x_{k+1}} \sum_{x_{k+2}} \dots \sum_{x_n} p(x_1, x_2, \dots x_n)$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей. Теорема Байеса

Совместное применение правила суммирования и обращения условного распределения позволяет получить выражение для **теоремы Байеса**:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\int p(x|y)p(y)dy}$$

- $p(y)$ – **априорное** распределение (лат. ***a priori*** – *из предшествующего*).
  ✓ Определяет наше изначальное знание о величине Y.
- $p(y|x)$ – **апостериорное** распределение (***a posteriori*** – *из последующего*).
  ✓ Определяет знание о Y после того, как мы пронаблюдали величину X.
- $p(x|y)$ – **правдоподобие** (*likelihood*).
  ✓ Определяет известный нам закон влияния величины Y на величину X.
- $p(x)$ – **обусловленность** (*evidence*).
  ✓ Нормировочная константа, безусловная вероятность величины X.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Основные понятия теории вероятностей. Условная зависимость и независимость

Две случайные величины X и Y **условно независимы** относительно третьей случайной величины Z ($X \perp\!\!\!\perp Y|Z$), тогда и только тогда, когда распределения их условных вероятностей относительно Z являются независимыми.

При условной независимости X и Y для каждого данного численного значения Z распределение вероятностей X не зависит от значений Y, и распределение вероятностей Y не зависит от значений X.

В случае условной независимости величин справедливо соотношение:

$$p(x, y|z) = p(x|z) \cdot p(y|z)$$

Если соотношение не выполняется, данные случайные величины называются **условно зависимыми** по Z.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Сущность байесовского моделирования

- При моделировании **все** величины считаются *случайными* и задаются вероятностными распределениями, в отличие от классических моделей, связывающих конкретные *значения* величин.
- Зависимость между случайными величинами задается в виде совместного распределения вероятностей.
- Байесовская модель, заданная совместным распределением вероятностей, позволяет определить **любые** маргинальные и условные вероятности всех величин, входящих в неё.
- При появлении новой информации имеется возможность определить вероятности связанных друг с другом событий с использованием теоремы Байеса.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

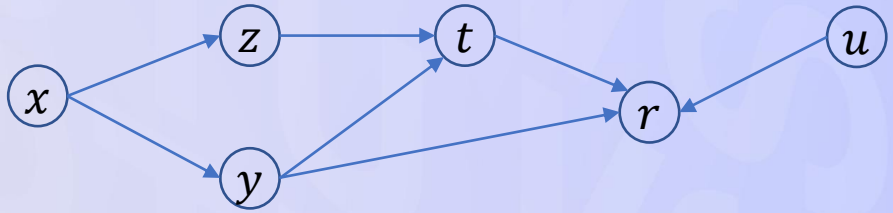
Графические вероятностные модели. Определение DAG

- Обычно, совместное распределение вероятностей включает в себя условные и безусловные независимости, которые могут упростить запись.
- Зависимости между переменными в байесовской модели часто визуализируют в форме направленного ациклического графа (DAG).
- **Узлы** графа содержат вероятностные распределения на случайные величины. Количество узлов графа равно числу случайных величин в совместном распределении вероятностей.
- **Рёбра** графа соответствуют зависимостям одной величины от другой.
- Если узел графа не содержит входящих ребер, то соответствующая ему случайная величина описывается безусловным распределением.
- Каждое входящее ребро в узел графа определяет величину, расположенную в родительском узле, по которой выражается условное распределение.

- Определенный таким образом граф называется байесовской сетью доверия.
- Он позволяет представить сложное совместное распределение в виде произведения более простых условных и безусловных распределений в соответствии со структурой графа:

$$p(x_1, x_2, \dots x_n) = \prod_{i=1}^{n} p(x_i \mid \mathrm{Parent}(x_i))$$

- Например:

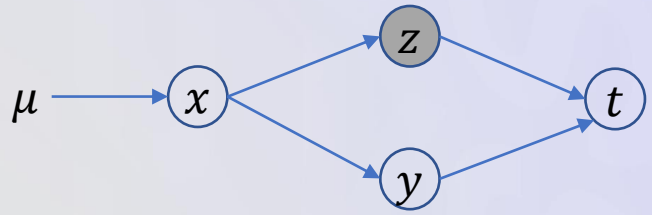


$$p(x, y, z, t, r, u) = p(x)p(z|x)p(y|x)p(t|y, z)p(r|u, y, t)p(u)$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Графические вероятностные модели. Наблюдаемые и латентные величины

- **Наблюдаемыми** (*observed*) называются величины, которые можно непосредственно измерить в ходе эксперимента. Узлы графа, соответствующие наблюдаемым переменным, обозначаются окрашенным кружком.
- **Латентные** (или скрытые, *latent*) переменные нельзя измерить напрямую; они могут быть вычислены косвенно с использованием наблюдаемых переменных. Узлы графа обозначаются неокрашенным кружком.
- **Параметры** (*parameters*) – это величины, представленные в вероятностной модели как некоторое фиксированное значение, а НЕ распределение.



$$p(x, y, z, t) = p(x, \mu)p(z|x)p(y|x)p(t|y, z)$$

$$p(x) = \mathcal{N}(\mu, 10) \qquad p(z|x) = \mathcal{G}(x, 1) \qquad p(t|y, z) = \mathcal{W}(y, 2z)$$

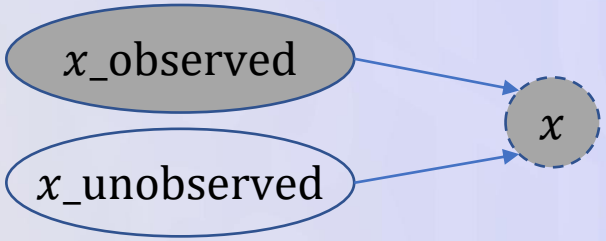
$$\mu \in [0, 5] \qquad p(y|x) = \mathcal{U}(0, x)$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Графические вероятностные модели. Частично-наблюдаемые величины

- Иногда конкретное значение некоторой случайной переменной в одном эксперименте может быть *известным* (переменная является наблюдаемой), а в другом – *отсутствовать* (случай латентной переменной).
- Такая переменная называется **частично-наблюдаемой** (*partial observed*).
- *Частично-наблюдаемые* переменные могут обозначаться несколькими способами:

- Иногда частично наблюдаемую переменную представляют как *комбинацию латентной* и *наблюдаемой* – в зависимости от того, задано ли явное её значение:



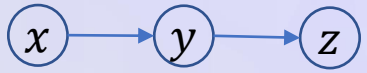
$$p(x) = \delta(x_{observed} \mid x_{unobserved})$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Графические вероятностные модели. d-разделимость

- Если в графической вероятностной модели некоторые величины являются наблюдаемыми, то существует несколько практических случаев, позволяющих уменьшить сложность графа.
- При появлении в вероятностной модели наблюдаемой переменной другие связанные с ней величины могут стать условно зависимыми или условно независимыми.
- **d-разделение** – это критерий, предназначенный для определения условной независимости двух величин при наличии третьей.
- Идея заключается в том, чтобы связать «*зависимость*» со «*связностью*» в графе и «*независимость*» с «*несвязностью*» или «*разделением*».

- **Последовательная связь** (chains) предполагает зависимость вида $X \to Y \to Z$
  - ✓ Случай, когда Y – латентная величина:

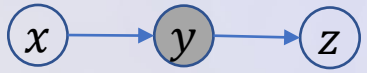


$$p(x,y,z) = p(x)p(y|x)p(z|y)$$

$$p(x,z) = p(x)\int p(y|x)p(z|y)dy = p(x)\int p(z|y,x)p(y|x)dy = p(x)p(z|x)$$

Переменные x и z зависимы.

  - ✓ Случай, когда Y – наблюдаемая величина:



$$p(x,y,z) = p(x)p(y|x)p(z|y)$$

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)p(y|x)p(z|y)}{p(y)} = \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y)$$

Переменные x и z условно независимы.

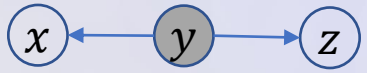
- **Расходящаяся связь** (forks) предполагает зависимость вида $X \leftarrow Y \rightarrow Z$
  - ✓ Случай, когда Y – латентная величина:

$$p(x,y,z) = p(y)p(x|y)p(z|y)$$

$$p(x,z) = \int p(y)p(x|y)p(z|y)dy$$

Переменные x и z зависимы.

  - ✓ Случай, когда Y – наблюдаемая величина:

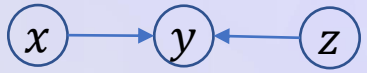


$$p(x,y,z) = p(y)p(x|y)p(z|y)$$

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y)$$

Переменные x и z условно независимы.

- **Сходящаяся связь** (v-structure) предполагает зависимость вида $X \to Y \leftarrow Z$
  - ✓ Случай, когда Y – латентная величина:

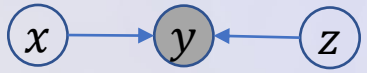


$$p(x,y,z) = p(x)p(z)p(y|x,z)$$

$$p(x,z) = p(x)p(z)\int p(y|x,z)dy = p(x)p(z)$$

Переменные x и z независимы.

  - ✓ Случай, когда Y – наблюдаемая величина:



$$p(x,y,z) = p(x)p(z)p(y|x,z)$$

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)p(z)p(y|x,z)}{p(y)}$$

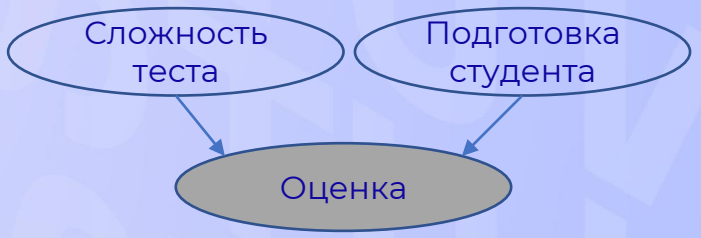
Переменные x и z условно зависимы!

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Графические вероятностные модели. Сходящаяся связь

- **Сходящаяся связь** представляет пример появления условной зависимости между *изначально независимыми* случайными величинами, при появлении наблюдаемой третьей случайной величины, зависящей от них.
- Данный эффект носит название «*explaining away*».
- Например:

О подготовке студента ничего нельзя сказать, зная сложность теста, и наоборот.

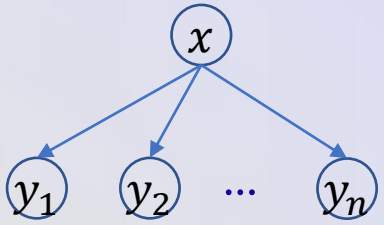


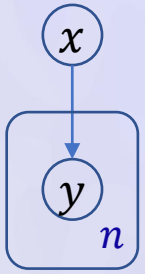
Сложность теста и подготовка студента стали зависимыми величинами. Если тест легкий, а оценка низкая, значит, студент недостаточно хорошо подготовился.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

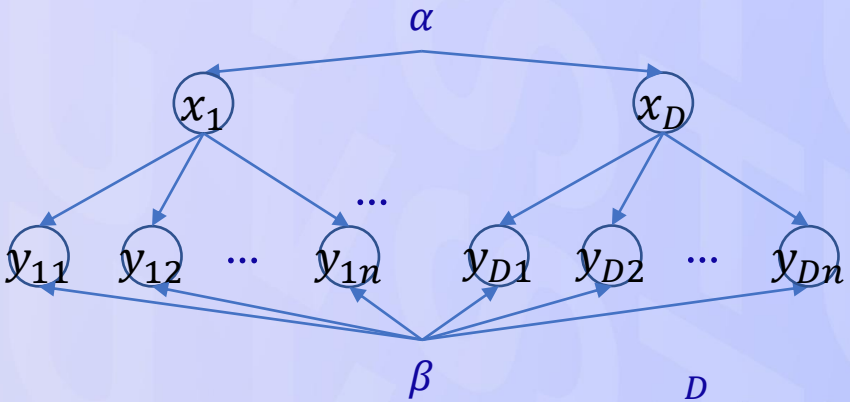
Графические вероятностные модели. Plate-нотация

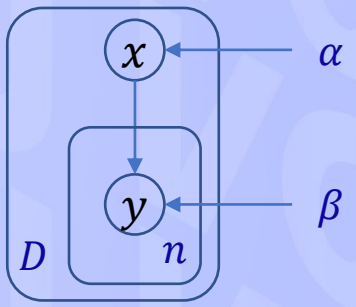
- Часто в графических моделях можно встретить множество похожих переменных с одинаковыми распределениями.
- При построении байесовской сети их объединяют в «**планки**» (*plates*, обозначаются как прямоугольники с указанием числа переменных), которые обозначают совместное распределение условно-независимых переменных.
- Например:





$$p(x, y_1, y_2, \dots y_n) = p(x)\prod_{i=1}^{n} p(y_i|x)$$





$$p(\alpha, \beta, x_1, x_2, \dots, y_1, y_2, \dots y_n) = \prod_{i=1}^{D} p(x_i|\alpha)\prod_{j=1}^{n} p(y_{ij}|x_i, \beta)$$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Графические вероятностные модели. Пример использования

- Пусть известны вероятностные зависимости об успехах изучения студентами некоторой дисциплины, представленные в виде байесовской сети.
  - ✓ *Оценка студента зависит от его интеллекта и сложности курса.*
  - ✓ *Студент может получить от преподавателя плохую или хорошую рекомендацию в зависимости от оценки студента.*
  - ✓ *Также студент сдаёт госэкзамен, результаты экзамена не зависят от рекомендации преподавателя, оценки за его курс и сложности курса.*
- Требуется вычислить:
  - ✓ *Вероятность получить хорошую рекомендацию студенту с низким интеллектом, если известно, что курс был легким.*
  - ✓ *Вероятность, что курс сложный, если студент получил оценку “C”.*

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Графические вероятностные модели. Пример использования

1. Определить вероятность получить *хорошую рекомендацию* студенту с *низким интеллектом*, если известно, что курс был *легким*.

   Полное распределение вероятностей задается выражением:

$$P(d,i,g,s,l) = P(d)P(i)P(g|d,i)P(s|i)P(l|g)$$

   Условную вероятность можно получить по формуле *условного распределения*, а затем выполнить *маргинализацию* по оставшимся переменным:

$$P(l_1|i_0,d_0) = \frac{P(l_1,i_0,d_0)}{P(i_0,d_0)} = \frac{\sum_g \sum_s P(d_0,i_0,g,s,l_1)}{\sum_l \sum_g \sum_s P(d_0,i_0,g,s,l)}.$$

Подставим в числитель и знаменатель *полное распределение* и преобразуем полученные выражения:

$$\sum_g \sum_s P(d_0, i_0, g, s, l_1) = \sum_g \sum_s P(d_0)P(i_0)P(g|d_0, i_0)P(s|i_0)P(l_1|g)$$

$$= P(d_0)P(i_0) \sum_g P(g|d_0, i_0)P(l_1|g) \sum_s P(s|i_0) = P(d_0)P(i_0) \sum_g P(g|d_0, i_0)P(l_1|g)$$

$$\sum_l \sum_g \sum_s P(d_0, i_0, g, s, l) = \sum_l \sum_g \sum_s P(d_0)P(i_0)P(g|d_0, i_0)P(s|i_0)P(l|g)$$

$$= P(d_0)P(i_0) \sum_g P(g|d_0, i_0) \sum_l P(l|g) \sum_s P(s|i_0) = P(d_0)P(i_0) \sum_g P(g|d_0, i_0) = P(d_0)P(i_0)$$

Таким образом, итоговое выражение примет вид:

$$P(l_1|i_0,d_0)=\frac{P(d_0)P(i_0)\sum_g P(g|d_0,i_0)P(l_1|g)}{P(d_0)P(i_0)}=\sum_g P(g|d_0,i_0)P(l_1|g)$$

Подставляя в него численные значения вероятностей, получаем:

$$P(l_1|i_0,d_0)=$$

$$P(g_A|d_0,i_0)P(l_1|g_A)+P(g_B|d_0,i_0)P(l_1|g_B)+P(g_C|d_0,i_0)P(l_1|g_C)=$$

$$0{,}3\cdot 0{,}9+0{,}4\cdot 0{,}6+0{,}3\cdot 0{,}01=0{,}513$$

Ответ: $P(l_1|i_0,d_0)=0{,}513.$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

1. Определить *вероятность, что курс сложный*, если студент получил *оценку "C"*.

   Полное распределение вероятностей задается выражением:

$$P(d,i,g,s,l) = P(d)P(i)P(g|d,i)P(s|i)P(l|g)$$

   Вероятность $P(d_1|g_C)$ можно рассчитать по *формуле обращения*:

$$P(d_1|g_C) = \frac{P(g_C|d_1)P(d_1)}{\sum_d P(g_C|d)P(d)} = \frac{P(g_C|d_1)P(d_1)}{P(g_C|d_0)P(d_0) + P(g_C|d_1)P(d_1)}$$

   Вероятности $P(g_C|d_1)$ и $P(g_C|d_0)$ нам *неизвестны*, то зато *известно* $P(g_C|d,i)$. Получим из него требуемые вероятности по *формуле суммирования*:

$$P(g_C|d_1) = \sum_i P(g_C|d_1, i)P(i) = P(g_C|d_1, i_0)P(i_0) + P(g_C|d_1, i_1)P(i_1) = 0{,}7 \cdot 0{,}7 + 0{,}2 \cdot 0{,}3 = 0{,}55$$

$$P(g_C|d_0) = \sum_i P(g_C|d_0, i)P(i) = P(g_C|d_0, i_0)P(i_0) + P(g_C|d_0, i_1)P(i_1) = 0{,}3 \cdot 0{,}7 + 0{,}02 \cdot 0{,}3 = 0{,}216$$

И итоговое выражение запишется в виде:

$$P(d_1|g_C) = \frac{P(g_C|d_1)P(d_1)}{P(g_C|d_0)P(d_0) + P(g_C|d_1)P(d_1)} = \frac{0{,}55 \cdot 0{,}4}{0{,}216 \cdot 0{,}6 + 0{,}55 \cdot 0{,}4} = 0{,}63$$

Ответ: $P(d_1|g_C) = 0{,}63.$

VOLGOGRAD STATE TECHNICAL UNIVERSITY

- **Тензор** (tensor) – *многомерный массив* чисел, основной объект данных в PyTorch. Является обобщением скаляров (0-мерный тензор), векторов (1-мерный тензор) и матриц (2-мерный тензор).
- Некоторые способы *создания* тензоров:
  - ✓ **`torch.tensor([1, 2, 3])`**: из списка Python;
  - ✓ **`torch.zeros(2, 3)`**: тензор, заполненный нулями:
  - ✓ **`torch.ones(5)`**: тензор, заполненный единицами.
- К основным *атрибутам тензора* относятся:
  - ✓ **Размерность** (ndim): Количество измерений тензора (количество осей) – **`tensor.ndim`**
  - ✓ **Форма** (shape): Кортеж, указывающий размер тензора по каждой оси – **`tensor.shape`**
  - ✓ **Тип данных** (dtype): Тип элементов тензора (torch.float32, torch.int64) – **`tensor.dtype`**
  - ✓ **Устройство** (device): Устройство, на котором находится тензор (CPU, GPU) – **`tensor.device`**

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Broadcasting в PyTorch. Правила broadcasting

- **Broadcasting** – это механизм в PyTorch, позволяющий выполнять арифметические операции над тензорами разных форм, если эти формы совместимы.
- Правила broadcasting:
  1. *Добавление измерений*: если у тензоров разная размерность, к форме меньшего тензора добавляются единицы слева, пока количество измерений не станет одинаковым.
  2. *Проверка совместимости размерностей*: два измерения считаются совместимыми, если:
     - ✓ они равны – в этом случае broadcasting вдоль этого измерения **не нужен**;
     - ✓ одно из них равно 1.

     Если измерения *несовместимы*, broadcasting **невозможен**.
  3. *Broadcasting*: Тензор, измерение которого равно 1, «**копируется**» вдоль этого измерения, чтобы соответствовать другому тензору.
  4. *Выполнение арифметической операции*: как только broadcasting сделает формы тензоров одинаковыми, выполняется арифметическая операция.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Broadcasting в PyTorch. Пример broadcasting

- Рассмотрим пример broadcasting на примере следующей операции:

```
a = torch.Tensor([[1], [2]])          # shape = (2, 1)
b = torch.Tensor([[[3, 4]]])          # shape = (1, 1, 2)
c = a + b                             # формы тензоров не совпадают.
```

- *Выровняем размерности* тензоров, добавив измерение слева к тензору **a**:

```
a = torch.Tensor([[[1], [2]]])        # shape = (1, 2, 1)
```

- *Выполним broadcasting* вдоль последнего и предпоследнего измерений у тензоров **a** и **b** соответственно:

```
a = torch.Tensor([[[1, 1], [2, 2]]])  # shape = (1, 2, 2)
b = torch.Tensor([[[3, 4], [3, 4]]])  # shape = (1, 2, 2)
```

- Формы тензоров совпали. *Выполняем операцию*:

```
c = torch.Tensor([[[4, 5], [5, 6]]])
```

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Распределения в PyTorch. Ключевые концепции

- **`torch.distributions`**: модуль PyTorch, предоставляющий классы для работы с различными *вероятностными распределениями*.
- Ключевые *концепции распределений* в PyTorch:
  - ✓ **Batch Shape**: Форма, соответствующая количеству *условно-независимых экземпляров* распределения, которые могут иметь *различные параметры*.
  - ✓ **Event Shape**: Форма *одного наблюдения* из распределения. Случайные величины, которые относятся к одному наблюдению, считаются *зависимыми*.
  - ✓ **Sample Shape**: Форма, указывающая сколько *независимых* и *одинаково распределённых* (i.i.d) наблюдений (объём и форму выборки) требуется получить из распределения.
  - ✓ **`sample`**(`sample_shape`): Генерирует *выборку* из распределения заданной формы sample_shape. Форма возвращаемого тензора: **`sample_shape + batch_shape + event_shape`**.
  - ✓ **`log_prob`**(`value`): Вычисляет *логарифм плотности вероятности* (или *вероятности* для дискретных распределений) для заданного значения `value`. `Value` должен иметь форму, совместимую с **`batch_shape`** и **`event_shape`**. *Результат* будет иметь размер **`batch_shape`**.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Распределения в PyTorch. Формирование форм распределений

- Каждое распределение в torch.distributions формирует *batch shape* и *event shape* в зависимости от типа распределения и формы параметров:
  - ✓ **`event_shape`**: Это форма, определяемая *внутренним свойством распределения*, определяющая размерность *одного наблюдения*. Оно может зависеть от размерности некоторых *параметров* (как в `MultivariateNormal` или `Dirichlet`), но сама зависимость *заложена в определение* самого распределения. Например:
    - *Одномерные распределения* (`Normal`, `Bernoulli`, `Exponential`): **`event_shape = ()`**
    - *Многомерные распределения* (`MultivariateNormal`, `Dirichlet`): **`event_shape`** будет равным размерности случайного вектора.
    - *Дискретные распределения*, принимающие значения из множества размера K (например, `Categorical`): **`event_shape = ()`**
  - ✓ **`batch_shape`**: Это форма, описывающая *независимые экземпляры распределения*. Она получается из формы параметров после того, как `event_shape` уже *определена* (и «изъята» из формы *параметров*, если это необходимо). Если параметры имеют разные формы, они приводятся к общей форме за счет *операции broadcasting*.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

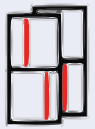
Распределения в PyTorch. Управление формами распределения

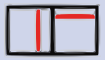
```
Normal(torch.tensor([0.0, 2.0]),          # batch_shape = (2,)
       torch.tensor([1.0, 1.0]))          # event_shape = ()
MultivariateNormal(torch.zeros(3, 2),     # batch_shape = (3,)
       torch.eye(2))                      # event_shape = (2,)
```

- Однако, иногда бывает полезно *переопределить формы распределения*.
- Для этого используются следующие *инструменты* манипуляции формами:
  - **`expand(batch_shape)`**: *Метод* распределения, возвращает новый *экземпляр* с формой батча, расширенной до `batch_shape`. Он вызывает метод `expand` параметров распределения. При этом новая память для сформированного экземпляра **не выделяется**.
  - **`Independent`**: Распределение, используется для переинтерпретации некоторых измерений `batch_shape` распределения как измерений `event_shape`. Параметры:
    - ✓ `base_distribution`: исходное распределение с переинтерпретируемыми формами;
    - ✓ `reinterpreted_batch_ndims`: количество крайних справа измерений `batch_shape` исходного распределения, которые нужно перенести в `event_shape`.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Распределения в PyTorch. Примеры форм распределений

Распределение	Сэмпл	Event shape	Batch shape	Sample shape	Описание
		()	()	()	Один сэмпл из одного одномерного распределения
		()	()	(2,)	Два сэмпла из одного одномерного распределения
		()	(2,)	()	Один сэмпл из двух одномерных распределений
		()	()	()	Два сэмпла из двух одномерных распределений

Распределение	Сэмпл	Event shape	Batch shape	Sample shape	Описание
		(2,)	()	()	Один сэмпл из одного двумерного распределения
		(2,)	()	(2,)	Два сэмпла из одного двумерного распределения
		(2,)	(2,)	()	Один сэмпл из двух двумерных распределений
		(2,)	(2,)	(2,)	Два сэмпла из двух двумерных распределений

https://ericmjl.github.io/blog/2019/5/29/reasoning-about-shapes-and-probability-distributions/

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Демонстрация практических примеров

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Заключение

1. Вспомнили основные сведения из высшей математики. Познакомились с байесовским моделированием, разобрались с основными принципами, лежащими в его основе.
2. Ввели понятие графических вероятностных моделей, разобрались с принципами их построения.
3. Познакомились с байесовскими сетями. Дали определение d-разделимости, рассмотрели основные случаи её применения.
4. Рассмотрели применение нотации «планок» (*plate notation*).
5. Разобрали примеры расчёта вероятностей с помощью графической модели.
6. В теории и на практических примерах познакомились с тензорами и распределениями в PyTorch.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

Спасибо за внимание!

Волгоград 2025