

Руководство пользователя MBayes

Содержание

1. О платформе.....	3
1.1. Системные требования.....	3
1.2. Теория, термины и определения.....	4
1.3. Применения.....	4
1.3.1. Анализ рисков.....	4
1.3.2. Финансовое моделирование.....	5
1.3.3. Цифровые двойники.....	5
1.3.4. Предиктивное обслуживание и ТОиР.....	5
2. Пользовательский интерфейс.....	5
2.1. Панели.....	7
2.1.1. Верхняя панель.....	7
2.1.1. Создать Граф.....	10
2.1.2. Редактор Графа.....	11
2.1.3. Редактор узла.....	11
2.1.4. Запуск.....	13
2.1.5. Гистограмма.....	14
2.1.6. Рассеяния.....	15
2.1.7. Обучение.....	16
2.1.8. Распределения.....	17
2.1.9. Сервер.....	19
2.2. Узлы.....	20
2.2.1. Категориальные узлы.....	20
2.2.2. Непрерывные узлы.....	21
2.2.3. Дискретные узлы.....	22
2.2.4. Детерминированные узлы.....	23
3. Процесс работы в Mbayes.....	24
3.1. Загрузка и создание модели.....	24
3.2. Редактирование модели.....	26
3.2.1. Создать узел.....	26
3.2.2. Добавить связь.....	26
3.2.3. Задать или обучить распределение.....	27
3.2.4. Ввести данные.....	28
3.3. Запуск вычислений.....	29
3.4. Анализ результатов.....	31
3.5. Экспорт модели.....	34
3.6. Завершение работы.....	34
4. Синтаксис формул и детерминированных узлов.....	34
5. Установка, запуск и удаление приложения.....	36
6. Решение типовых проблем (FAQ).....	36
7. Контакты и техническая поддержка.....	37

1. О платформе

MBayes - платформа для разработки и проектирования Интерпретируемого ИИ на основе вероятностных графовых моделей представляет собой интегрированную систему создания и обучения статистических моделей и может использоваться в различных областях, где:

- необходимо работать со стохастическими величинами;
- важна интерпретируемость и надежность выводов;
- требуется оценка неопределенности значений параметров;
- моделируемые процессы имеют нелинейные зависимости, иерархическую структуру и скрытые переменные;
- могут отсутствовать большие объемы данных

Система представляет собой конструктор для иерархических статистических моделей (Байесовских сетей), создаваемых используя данные или/и мнения экспертов.

Цель системы - упрощение создания моделей и алгоритмов для вычислений на их основе, что позволяет экспертам в предметных областях сконцентрироваться на моделировании, повышая скорость разработки аналитических приложений в разы, снижая проектные риски и повышая интерпретируемость моделей.

Пользователи получают возможность применять сложные вероятностные методы без знания и необходимости программирования, что уменьшает количество ошибок в коде и сокращает временные и стоимостные затраты на реализацию отраслевых решений. Кроме этого, MBayes делает модели визуальными, что облегчает их понимание и обсуждение, а также уменьшает количество ошибок моделирования.

1.1. Системные требования

Для эффективной работы MBayes рекомендуется использовать устройства с следующей конфигурацией:

- Операционная система (ОС): Ubuntu Linux (≥ 22.04 LTS), Windows 10
- Браузер: Firefox (≥ 143) Yandex Browser (≥ 25)
- Docker:
 - Для Windows (Docker Desktop версия ≥ 4.34)
<https://docs.docker.com/desktop/setup/install/windows-install/>
 - Для Linux
<https://docs.docker.com/engine/install/>

Рекомендуемые системные требования для работы Mbayes

Параметр	Значение
Количество процессорных ядер	8
Оперативная память (ОЗУ)	16 Гб
Свободное место на диске	10 Гб

Платформа состоит из двух компонентов – клиентского и серверного приложений. В рамках данного документа будет рассмотрен сценарий локального запуска и работы системы. Для решения масштабных задач возможен запуск нескольких серверных приложений параллельно.

1.2. Теория, термины и определения

В рамках платформы, ‘модель’ состоит из множества узлов и связей формирующих направленный ациклических граф. Связи указывают направление влияния, а узлами обозначаются случайные переменных модели. Вместе модель является представлением многомерного совместно распределения:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i))$$

Для целей обработки информации и интеграции данных в модели создаются узлы для данных. Эти узлы определяют форму функции правдоподобия. При поступлении данных их значения добавляются в граф и позволяют алгоритмам, реализованным в MBayes, вычислить апостериорные распределения на основе теоремы Байеса для остальных узлов модели. Далее, для создания прогнозов, производится выборка из заданного распределения, что позволяет адаптировать прогноз при поступлении новой информации в модели.

1.3. Применения

Платформа имеет применения в областях, где требуется стохастическое моделирование процессов, интеграция данных и поддержка принятия решений в условиях неопределенности, а также оценка вероятности событий. Ниже представлено несколько возможных областей применения системы.

1.3.1. Анализ рисков

В рамках задач по анализу рисков и стандартов ИСО/ГОСТ Р 31010, платформа предоставляет

- Байесовский анализ и сети (Bayesian analysis & networks)
- Анализ причин и последствий (Cause consequence analysis)
- Анализ дерева неисправностей (Fault tree)
- Анализ дерева событий (Event tree)
- Марковский анализ (Markov Analysis)
- Моделирование методом Монте-Карло (Monte Carlo Simulation)

1.3.2. Финансовое моделирование

В рамках задач по финансовому моделированию система позволяет создавать стохастические модели различных классов рисков и их корреляций, проводить сценарный анализ, оценивать вероятности дефолта контрагентов, решать задачи в области оптимизации стратегий перестрахования (reinsurance optimization).

1.3.3. Цифровые двойники

В рамках задач по созданию цифровых двойников процессов платформа MBayes позволяет пользователям создавать иерархические модели процессов с учетом вариабельности параметров систем используя различные распределения. При вводе реальных данных система позволяет получать оценки значений параметров на основе Байесовского вывода.

1.3.4. Предиктивное обслуживание и ТОиР

В рамках задач по предиктивному обслуживанию платформа позволяет создать модели, способные прогнозировать остающийся срок полезного использования (RUL, remaining useable lifetime). В рамках ремонта и диагностики, MBayes можно применить для оптимизации процесса диагностики неисправностей через выбор оптимального порядка тестирования возможных поломок.

2. Пользовательский интерфейс

Приложение открывается на главной странице редактора представленной на рис 1.

Раздел  Граф в правой стороне отображает все узлы модели, которая загружена в систему. Пример модели с четырьмя узлами представлен на рис 2. Панель «Граф» можно закрыть при нажатии на кнопку  Граф в левой части экрана или на .

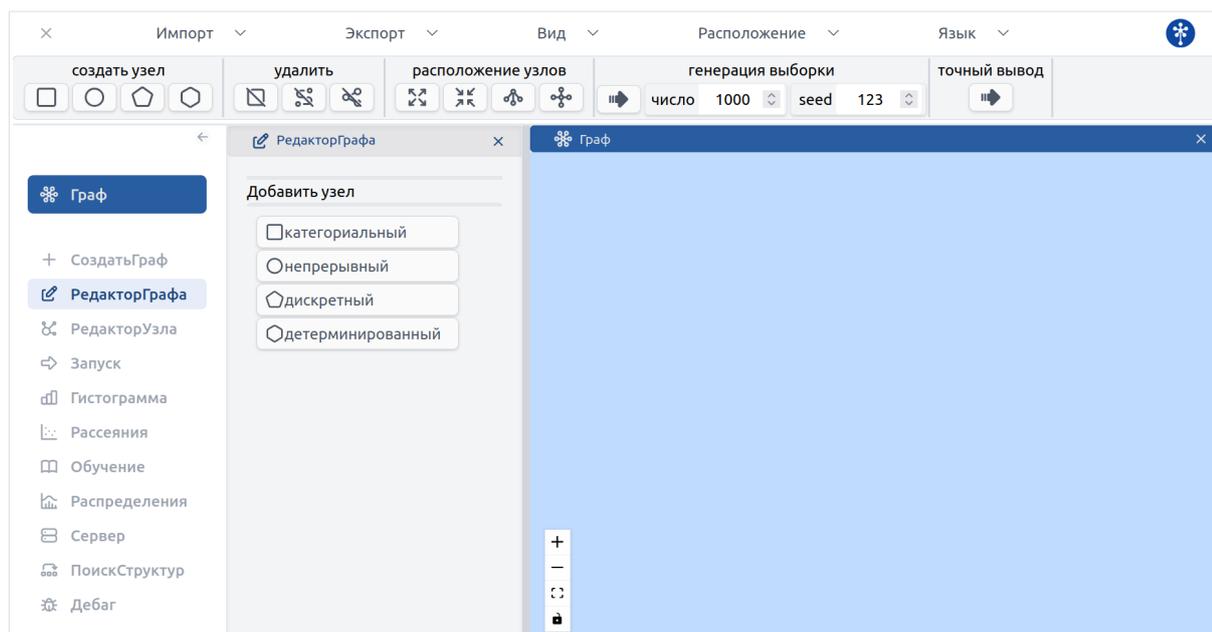


Рис. 1: Вид главой страницы приложения

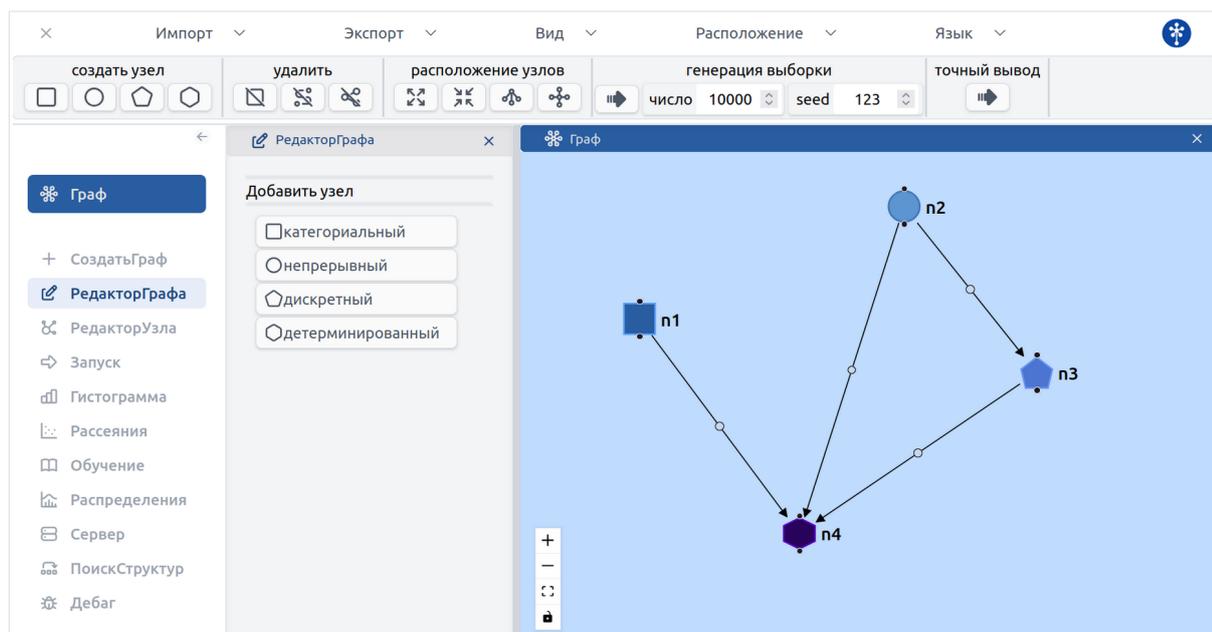


Рис. 2: Пример основного окна редактора с моделью состоящей из четырех узлов в свернутом виде

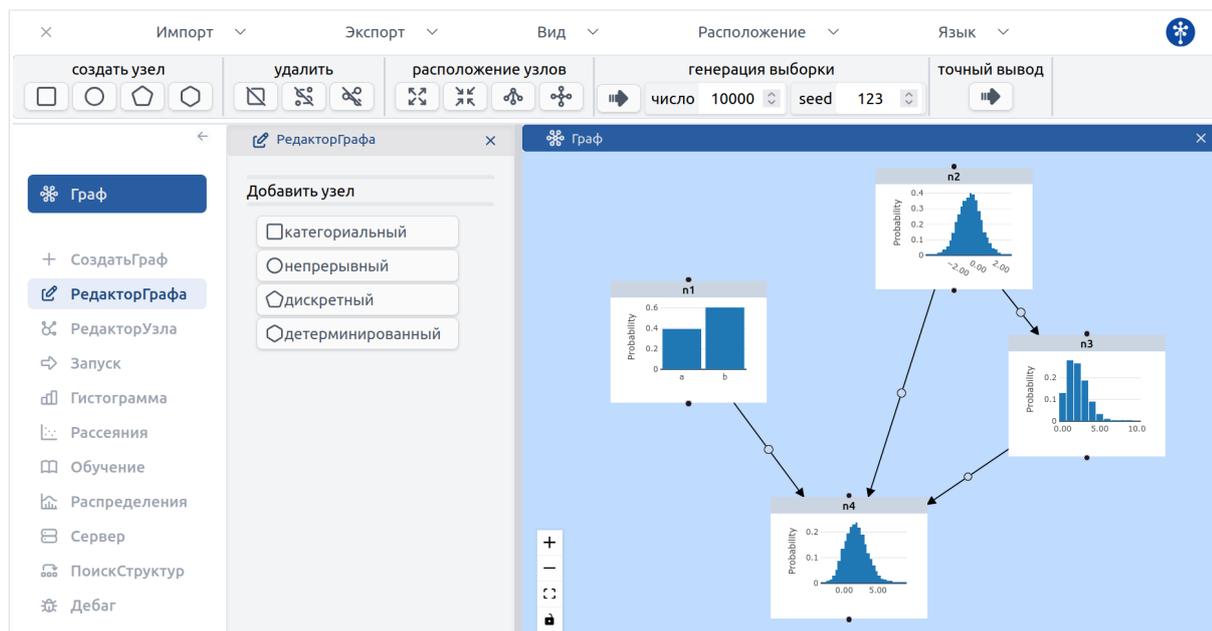


Рис. 3: Вид редактора с узлами модели представленными в 'развернутой' форме

2.1. Панели

Создание и редактирование моделей, а также ввод данных, запуск вычислений и анализ результатов осуществляется через различные панели, представленные в верхней и левой частях приложения. Данный раздел дает краткое описание каждой панели и ее функций.

2.1.1. Верхняя панель

Верхняя панель содержит функции для импорта и экспорта результатов, а также настройки параметров визуализации, выбора языка и общего вида приложения.

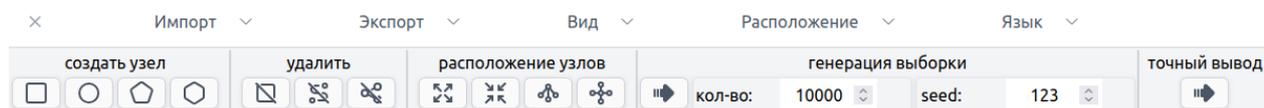


Рис. 4: Верхняя панель редактора

2.1.1.1. Меню

Содержит разделы: Импорт, Экспорт, Вид, Расположение, Язык.

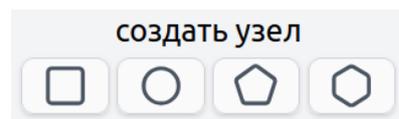
- **Импорт : для загрузки моделей в систему**
 - модель из JSON – загрузка модели из JSON файла
- **Экспорт : для экспорта моделей, выборки и визуализаций**
 - сэмплов – экспорт сгенерированной выборки в формате csv
 - маргинальных распределений – экспорт маргинальных распределений
 - модель – экспорт модели в формате JSON
 - модели и метаданные – экспорт модели и метаданных в формате JSON
 - png image – экспорт вида графа в формате PNG
- **Вид: настройка параметров визуализации**
 - гистограмма – определяет количество интервалов на числовой прямой
 - мини-карта – включает/выключает миникарту графа
 - название узлов – показать/скрыть название узлов и идентификатор узлов
 - edge handle – показать/скрыть символ удаления связей между узлами
 - коннектор узлов – показать/скрыть коннектор для создания связей между узлами
- **Расположение: запуск алгоритмов расстановки узлов модели**
 - развернуть – развернуть все узлы модели чтобы показать распределения
 - свернуть – свернуть все узлы модели чтобы показать иконки узлов
 - силовая расстановка – запустить алгоритм силовой расстановки узлов
 - dagre – запустить алгоритм расстановки узлов 'dagre'
 - ranksep – задает параметр 'ranksep' алгоритма 'dagre'
 - nodesep – задает параметр 'nodesep' алгоритма 'dagre'
 - edgesep – задает параметр 'edgesep' алгоритма 'dagre'
- **Язык: выбор языка**
 - English – выбор Английского языка
 - Русский – выбор Русского языка

2.1.1.2. Меню быстрого доступа

Содержит разделы:

Создать узел, Удалить, Расположение узлов, Генерация выборки, Точный вывод

- **Создать узел : создает новый узел в модели**



Категориальный узел



Непрерывный узел

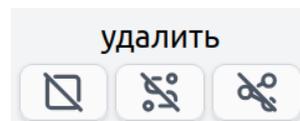


Дискретный узел



Детерминированный узел

- **Удалить**



Удалить узел

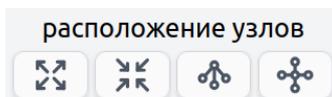


Удалить связь



Очистить граф

- **Расположение узлов: запуск авторасположения**



Категориальный узел



Непрерывный узел



Дискретный узел



Детерминированный узел

- **Генерация выборки:
запуск Монте-Карло**

генерация выборки

▶ число 1000 seed 123



Запускает вычисления

число 1000

Задаёт размер выборки

seed 123

Задаёт значение 'seed'

- **Точный вывод : запуск алгоритмов точного вывода**

ТОЧНЫЙ ВЫВОД



Запускает вычисления



2.1.1. Создать Граф

Загрузить модель

Содержит функционал для загрузки модели из JSON файла где по происхождению авто-расположение узлов (по умолчанию).

Несвязанный граф

Раздел позволяет провести автоматическую идентификацию узлов модели из CSV файла с данными. На каждую колонку файла будет создан один узел с автоматически определенным типом.

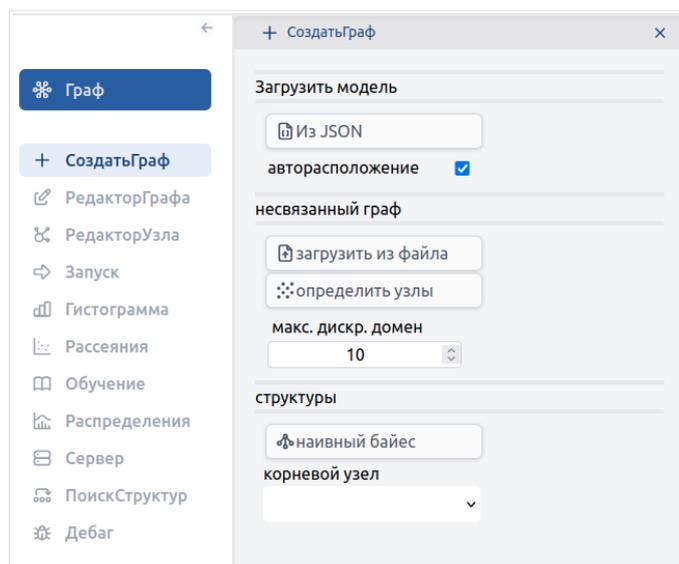


Рис. 5: Панель 'Создать граф'

Параметр 'макс. дискр. домен' обозначает максимальный домен для дискретных узлов, если же дискретный узел содержит больше уникальных значений, чем этот предел, то автоматическое определение создаст непрерывный узел.

Структуры

Раздел позволяет автоматически задать одну из стандартных структур модели. В демо версии реализована структура *Наивного Байесовского Классификатора* с возможностью выбора корневого узла.

2.1.2. Редактор Графа

Панель позволяет создать новые узлы в модели

категориальный

Создает стохастический категориальный узел

непрерывный

Создает стохастический непрерывный узел

дискретный

Создает стохастический дискретный узел

детерминированный

Создает детерминированный узел

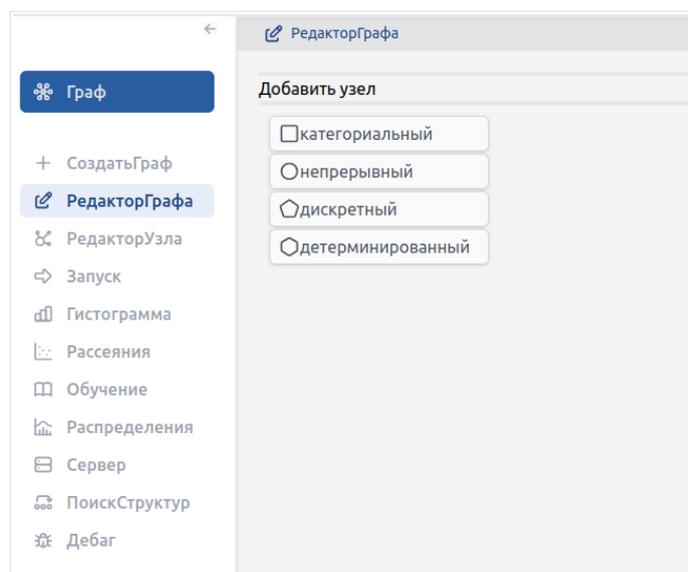
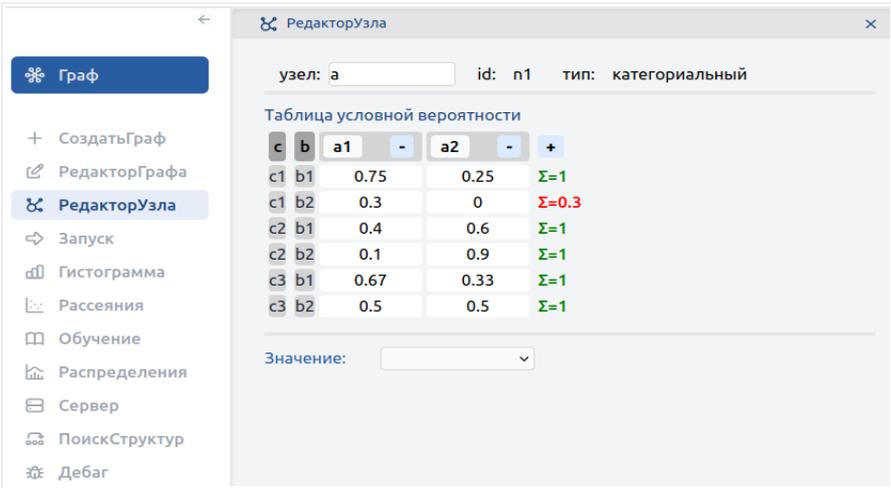


Рис. 6: Панель 'Редактор Графа'

2.1.3. Редактор узла

Панель позволяет редактировать априорные условные распределения узла как функции от узлов родителей. В панели отображен один из возможных форматов условной вероятности выбранного узла, которая зависит от типа узла. При выборе другого узла (двойным нажатии на соответствующий узел в панели 'Граф') эта панель обновляется автоматически. На рис. 7 представлена таблица условной вероятности для категориального узла.



c	b	a1	a2	
c1	b1	0.75	0.25	$\Sigma=1$
c1	b2	0.3	0	$\Sigma=0.3$
c2	b1	0.4	0.6	$\Sigma=1$
c2	b2	0.1	0.9	$\Sigma=1$
c3	b1	0.67	0.33	$\Sigma=1$
c3	b2	0.5	0.5	$\Sigma=1$

Рис. 7: Таблица условной вероятности узла

В примере на рис 7 колонками **c** и **b** таблицы обозначаются узлы 'родители' и показаны все возможные комбинации их значений, а колонками **a1** и **a2** обозначаются возможные значения, которые принимает величина узла a. Сумма вероятностей каждой строки должна быть равна единице ($\Sigma=1$), в противном случае система покажет ошибку, к примеру во второй строке ($\Sigma=0.3$). Добавление нового возможного значения происходит используя **+**, а удаление через **-**.

Для непрерывного узла таблица условной вероятности представлена на рис 8, в этом случае можно выбрать тип распределения и значения его параметров, а также форму функциональной зависимости от узлов родителей.

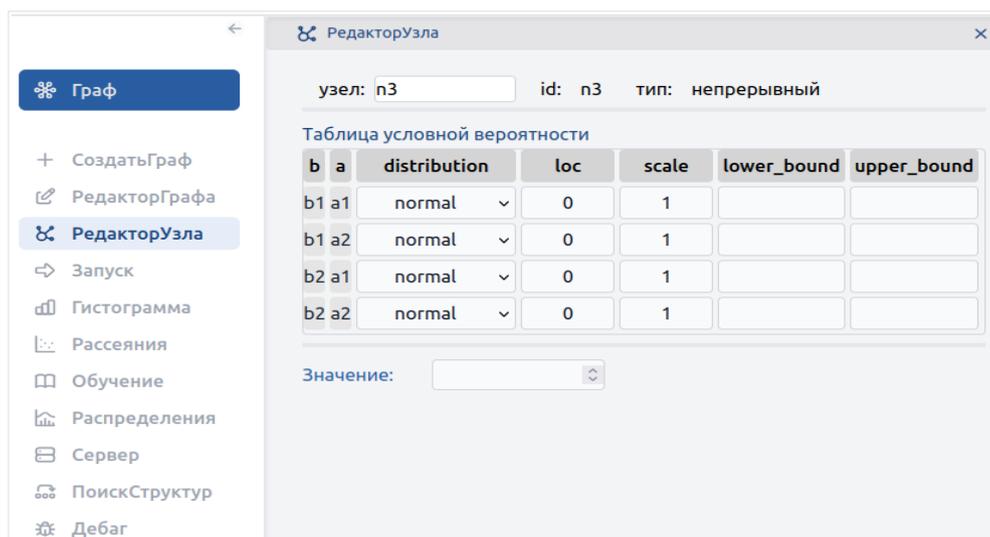


Рис. 8: Таблица условной вероятности \circ узла

В непрерывных узлах реализован следующий набор распределений. Использование значений 'lower_bound' и 'upper_bound' (нижняя / верхняя граница) **не обязательно**.

Идентификатор	Распределение
normal	Распределение Гаусса
inverse_gamma	Инверсное Гамма
uniform	Равномерное
cauchy	Распределение Коши
exponential	Экспоненциальное
laplace	Распределение Лапласа
beta	Бета
chi_square	Хи-квадрата
gamma	Гамма
logistic	Логистическое
student_t	Распределение Стьюдента
triangular	Треугольное
pareto	Распределение Парето

Значение

Ввод данных в узел происходит через выбор варианта (для категориального или дискретного типа узлов) или ввод значения (для непрерывного типа узлов) в поле 'значение'.

В дискретных узлах реализован следующий набор распределений:

Идентификатор	Распределение
poisson	Распределение Пуассона
binomial	Биномиальное распределение
bernoulli	Распределение Бернулли
negative_binomial	Отрицательное Биномиальное
randint	Дискретное равномерное

Использование значений 'lower_bound' (нижняя граница) и 'upper_bound' (верхняя граница) **обязательно**.

2.1.4. Запуск

Панель запуска предназначена для запуска вычислительных алгоритмов с целью определения априорных и апостериорных распределений, как одномерных, так и многомерных.

Генерация выборки

Задаёт число симуляций, которые нужно создать на основе метода Монте-Карло и инициализирующее значение (seed) для Генератора Случайных Чисел (ГСЧ). Метод применим для решения общего класса задач для непрерывных, дискретных и гибридных моделей. Однако полученные распределения и оценки их свойств являются приближенными, для повышения точности используйте более высокое значение параметра 'число'.

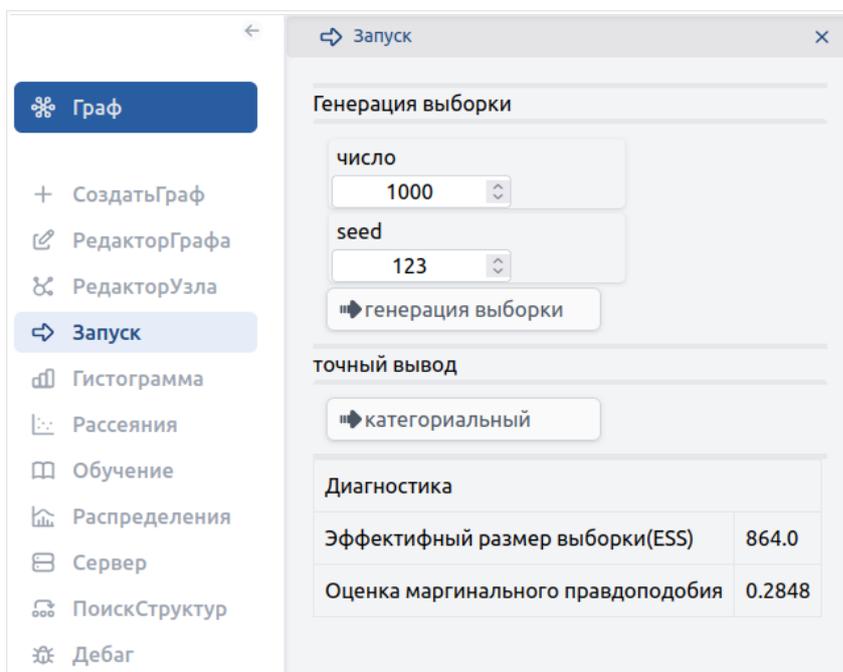


Рис. 9: Панель запуска вычислений (инференса)

Точный вывод

Запускает алгоритм точного вывода для графов, состоящих из категориальных узлов. Метод дает точные оценки априорных и апостериорных вероятностей.

Диагностика

При вводе данных в узлы модели раздел диагностика показывает диагностические параметры вывода алгоритма Монте-Карло:

- Эффективный размер выборки (effective sample size)
- Оценка маргинального правдоподобия (marginal likelihood estimate)

2.1.5. Гистограмма

При завершении работы вычислительных алгоритмов панель 'Гистограмма' может быть использована для визуализации одномерных распределений узлов модели и анализа параметров распределений. Панель показывает распределение активного узла, смены активного узла делается (быстрым) двойным нажатием мыши.

Логарифмическая шкала

Включает логарифмическую шкалу на оси у, по умолчанию шкала линейная.

Апостериорное значение

Включает отображение апостериорного распределения (при его наличии) в случае, когда данные введены хотя бы в один узел модели.

Функция распределения

Включает отображение функции распределения (CDF) случайной величины. По умолчанию показывает плотность распределения (PDF).

Гистограмма

Задает количество интервалов для гистограммы.

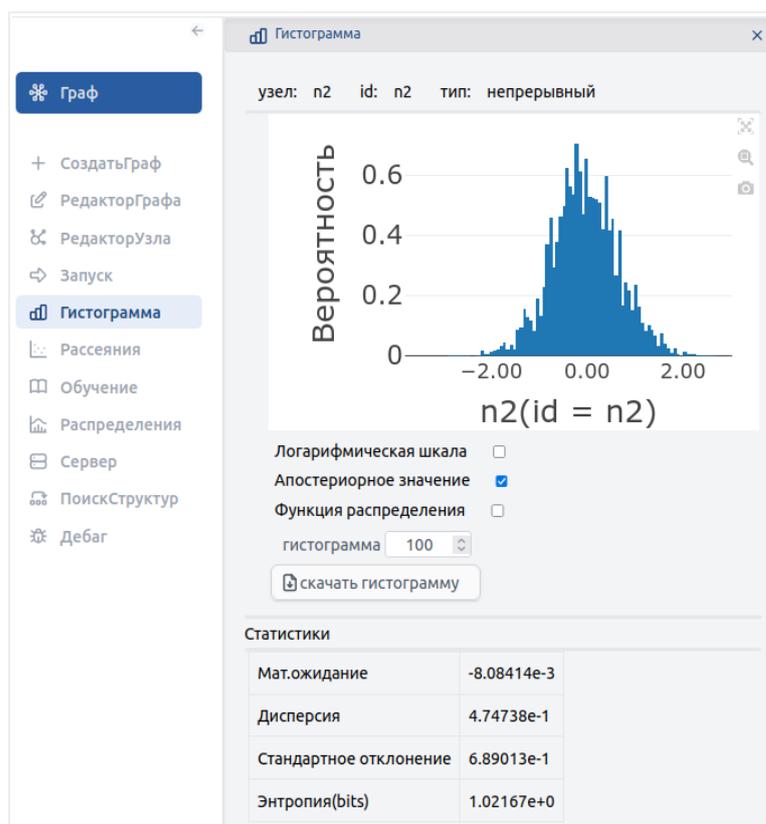


Рис. 10: Вид панели 'Гистограмма'

Статистики

Выводит значение следующих статистик распределения: *математическое ожидание, дисперсия, стандартное отклонение, энтропия.*

2.1.6. Рассеяния

Панель предназначена для визуализации и анализа двумерных совместных распределений, как априорных, так и апостериорных. После запуска вычислений требуется выбрать два узла, совместное распределение которых нужно отобразить через выбор идентификатора узлов в верхней части панели в разделах 'ось: X' и 'ось: Y', после чего приложение автоматически отобразит двумерное распределение и статистики.

Апостериорное значение

Включает отображение апостериорного распределения (при его наличии), в случае когда данные введены в (как минимум) один узел модели.

Гистограмма

Задаёт количество интервалов для двумерной гистограммы.

Статистики

Выводит значение следующих статистик распределения:

- *математическое ожидание*
- *ковариация*
- *взаимная информация*

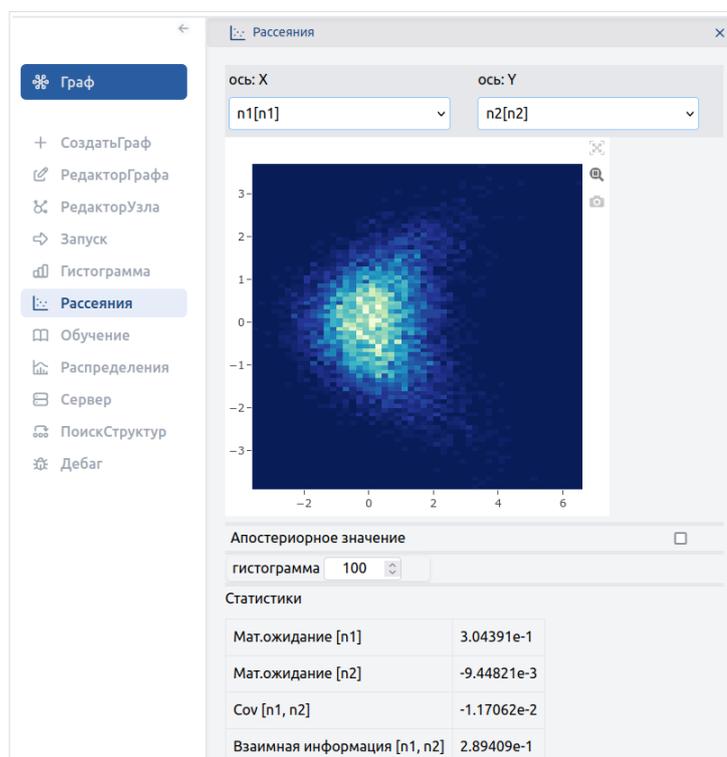


Рис. 11: Вид панели 'Рассеяния'

2.1.7. Обучение

Панель предназначена для обучения параметров условных распределений узлов в процессе первичного построения модели. Параметры и методы их калибровки зависят от типа узла и распределения в нем, а также желаемой функциональной формы зависимости узла от его родителей. В процессе обучения используется метод максимального правдоподобия. В случае калибровки непрерывных узлов, зависящих от непрерывных узлов родителей, автоматически задается линейная форма зависимости.

Обучение параметров

Позволяет загрузить данные для обучения всех узлов модели в форме CSV файла, где столбцы соответствуют идентификаторам узлов модели, а строки являются данными. В демо версии не поддерживается калибровка модели/узлов с частично отсутствующими данными.

Параметры алгоритмов обучения

Список параметров алгоритмов обучения. В демо версии доступен только параметр 'псевдосчетчики' для калибровки категориальных распределений, используя алгоритм 'laplace smoothing'.

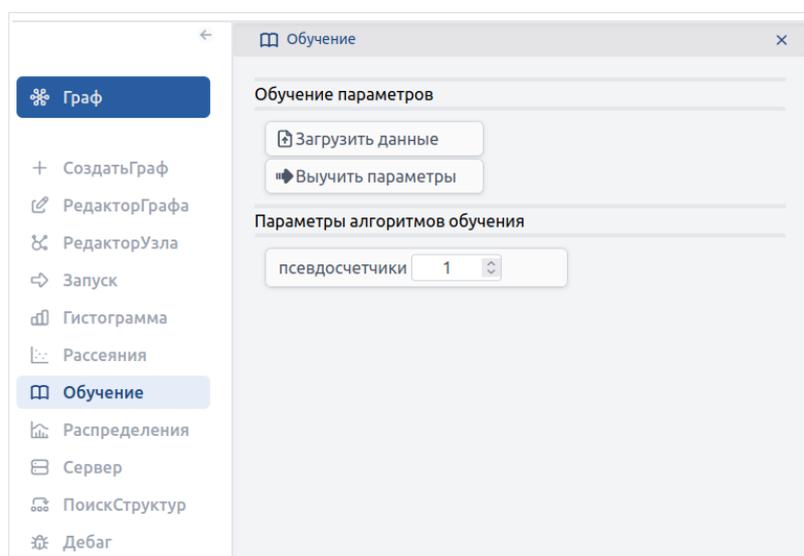


Рис. 12: Вид панели 'Обучение'

2.1.8. Распределения

Панель предназначена для анализа и подгонки распределений из данных, загруженных пользователем или из результатов симуляции.

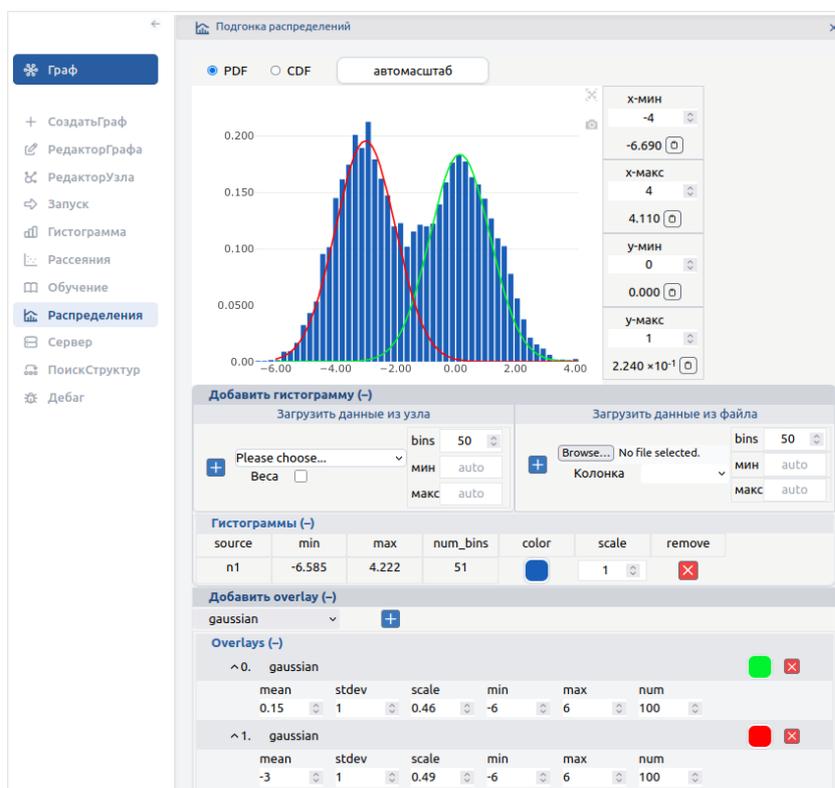


Рис. 13: Вид панели 'Распределения'

Шкалы осей графика и функция автомасштаб

В верхней части панели можно задать ручную шкалы осей X и Y для визуализации или, при нажатии 'автомасштаб' автоматически вычислить оптимальные шкалы осей для визуализации гистограмм и кривых распределения (overlays) на графике.

Добавить гистограмму

Позволяет добавить гистограмму на график, использует общие параметры:

- bins : количество интервалов гистограммы
- мин : минимальное значение интервала гистограммы
- макс : максимальное значение интервала гистограммы
- scale: определяет масштабирующий коэффициент, на который умножается высота блоков гистограммы.

Загрузить данные из узла: позволяет создать гистограмму из данных узла модели (при их наличии).

- *веса*: включает использование весов выборки при определении гистограммы

Загрузить данные из файла: позволяет создать гистограмму из данных, загруженных в приложение из CSV файла

- *колонка* : выбирает колонку файла, данные которой будут использоваться для создание гистограммы

Добавить overlay

Позволяет добавить кривые плотности распределений на график. После добавления каждая плотность контролируется через заданные параметры, а также общие параметры:

- *scale*: определяет масштабирующий коэффициент на который умножается значение плотности распределения
- *min* : минимальное значение интервала плотности
- *max* : максимальное значение интервала плотности
- *num* : количество точек на интервале [мин, макс]

2.1.9. Сервер

Панель предназначена для выбора сервера, на котором запущено серверное приложение. Выбор сервера делается нажатием на одну из конфигураций (на рис. 14 показана только одна конфигурация с названием 'localhost') или через ручной ввод url и порта сервера. В рамках демоверсии изменения списка конфигураций недоступны.

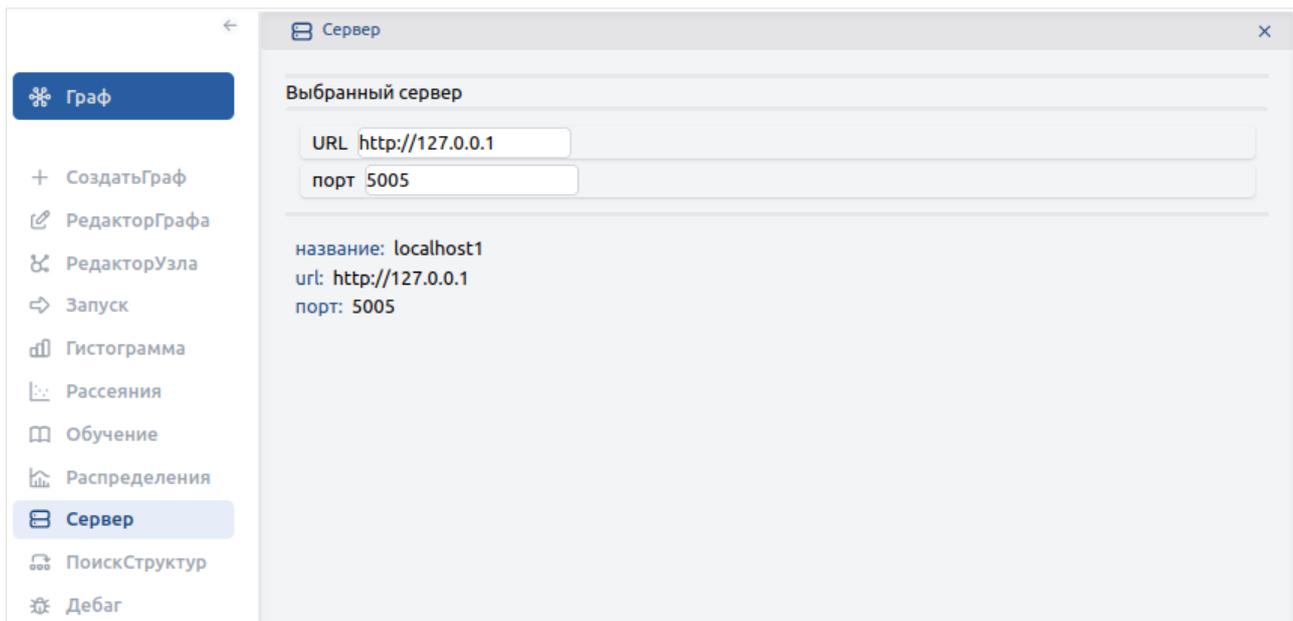
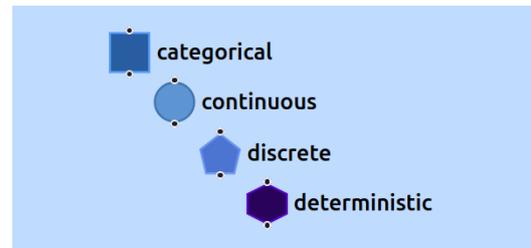


Рис. 14: Вид панели 'Сервер'

2.2. Узлы

Ключевым компонентом модели являются узлы, каждый из которых обозначает случайную или детерминированную величину. В рамках данной версии платформы MBayes можно работать с четырьмя основными типами узлов:

- категориальные (categorical)
- непрерывные (continuous)
- дискретные (discrete)
- детерминированные (deterministic)



Первые три типа узлов являются стохастическими, т.е. случайными переменными, в то время как детерминированный узел является нестохастической функцией, зависящей от узлов-родителей.

В рамках модели каждый узел имеет уникальный идентификатор (node_id), который используется для формирования зависимостей между узлами.

2.2.1. Категориальные узлы



Узлы используются для представления категориальных случайных переменных, имеющих форму:

$$X \sim \text{Categorical}(p_1, p_2, \dots, p_n)$$

В рамках модели узлами-родителями категориальных узлов могут выступать другие категориальные, дискретные или детерминированные узлы. Каждый категориальный узел может иметь различное количество состояний (n) и в случае ввода 'значений' принимает одно из них. В случае, когда категориальный узел имеет несколько дискретных узлов-родителей, приложение автоматически создает таблицу условных вероятностей, содержащую все комбинации родителей. Пример такой таблицы представлен на рис. 15, где показана таблица условных вероятностей узла 'n1', зависящих от узлов 'n2' с состояниями [a1,a2] и узла 'n3' с состояниями [b1,b2,b3].

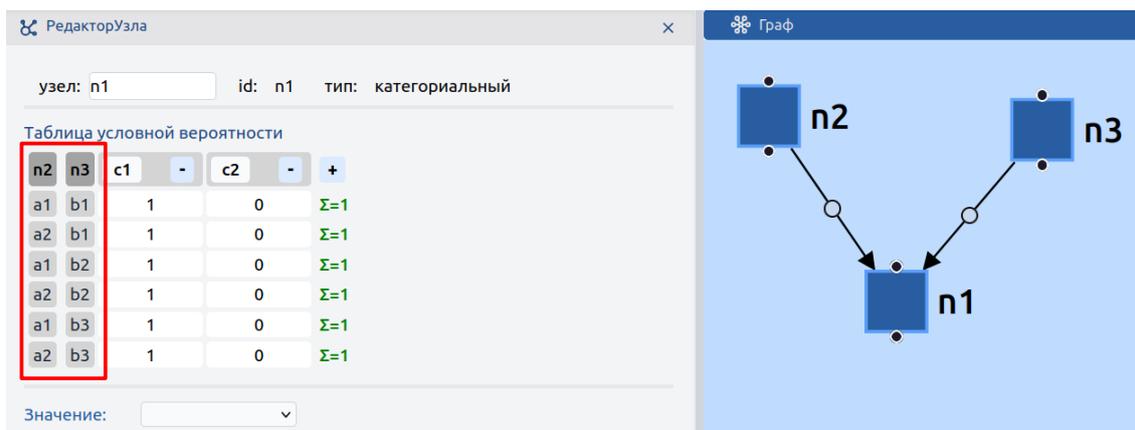


Рис. 15: Таблица условных вероятностей узла 'n1', имеющего узлы-родители 'n2' и 'n3' принимающие значения [a1,a2] и [b1, b2, b3].

2.2.2. Непрерывные узлы



Узлы используются для непрерывных случайных переменных. В рамках модели узлами-родителями непрерывных узлов могут выступать все типы узлов. В случае если узел-родитель непрерывный, зависимости прописываются в таблице условной вероятности. Пример линейной зависимости условного распределения узла n2 от узла n1 по параметру математического ожидания Гауссовской случайной величины показан в рис. 16.

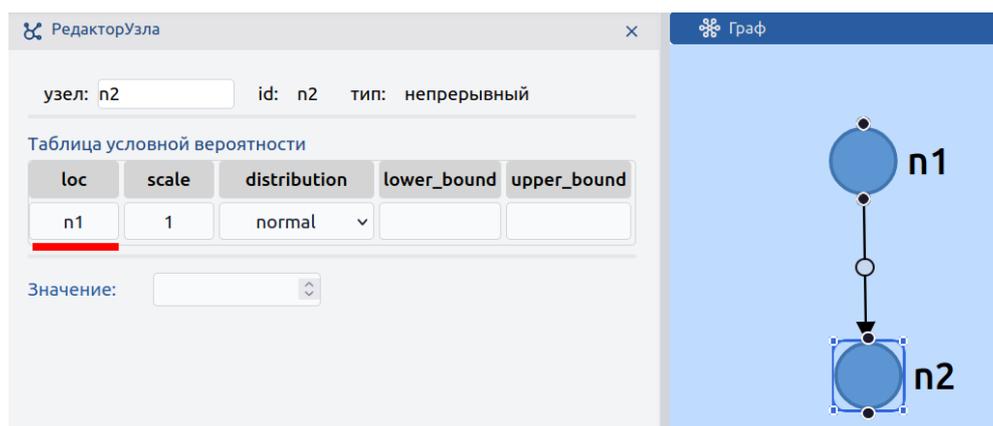
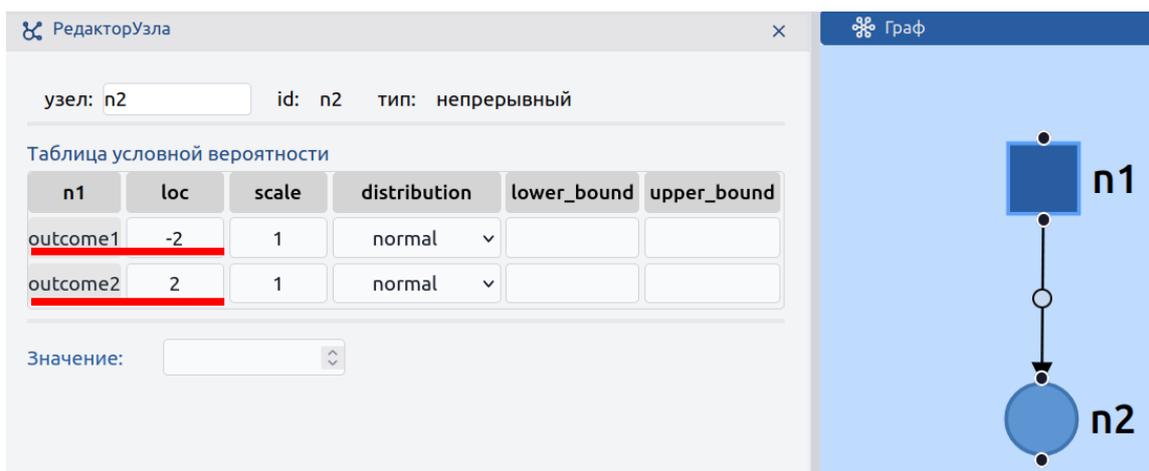


Рис. 16: Зависимости условной вероятности узла 'n2' в параметре 'loc' от узла родителя 'n1'

В случае, когда непрерывный узел имеет категориального/дискретного родителя,

приложение автоматически создает таблицу условных вероятностей, в рамках которой для каждого варианта возможно задать различные параметры условных распределений. В случае, рассмотренном на рис. 17 когда узел n1 имеет два состояния ('outcome1' и 'outcome2'), таблица условных вероятностей n2 содержит две строки – одну для каждого из вариантов, в каждый из которых можно ввести различные параметры распределений.



узел: n2 id: n2 тип: непрерывный

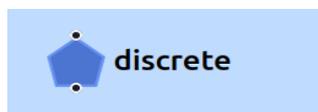
Таблица условной вероятности

n1	loc	scale	distribution	lower_bound	upper_bound
outcome1	-2	1	normal		
outcome2	2	1	normal		

Значение:

Рис. 17: Зависимости условной вероятности узла 'n2' от категориального узла-родителя 'n1'

2.2.3. Дискретные узлы



Используются для дискретных случайных переменных. В рамках модели, узлами-родителями дискретных узлов могут выступать все типы узлов. В случае если узел-родитель непрерывный, зависимости прописываются в таблице условной вероятности с учетом ограничений значений параметров используемых распределений. В отличие от непрерывных узлов, где нижние (lower bound) и верхние (upper bound) границы необязательны, в случае когда узел дискретный они требуются всегда. Пример случая, когда дискретный узел имеет непрерывного родителя, показан на рис. 18 где используется оператор абсолютного значения (abs), для того чтобы гарантировать неотрицательность параметра 'mu' (интенсивности) в Пуассоновском распределении, а интервал распределения ограничен отрезком [0,10].

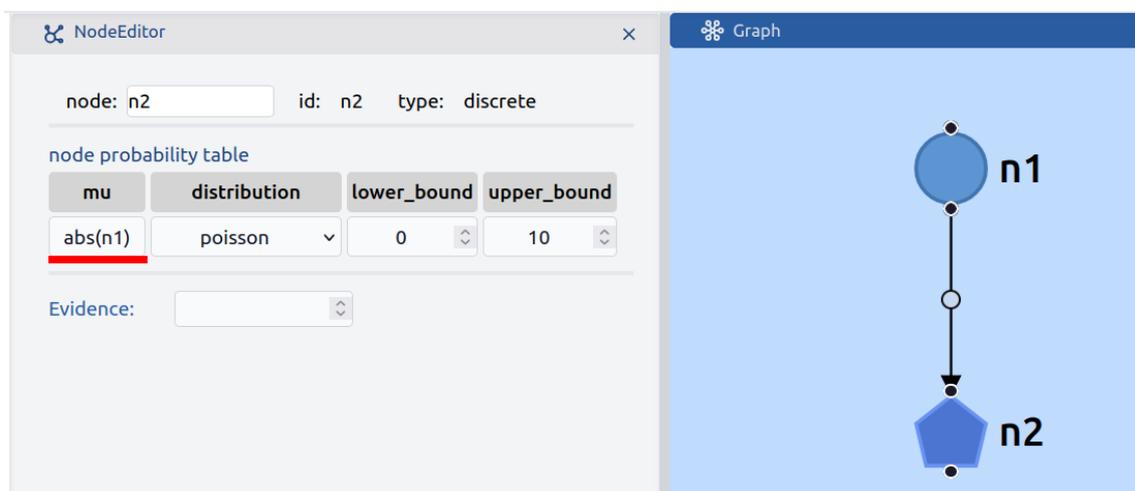


Рис. 18: Условная зависимость дискретного узла 'n2' с Пуассоновским распределением от непрерывного узла 'n1'.

2.2.4. Детерминированные узлы



Используются для представления детерминированных функций параметров. В рамках модели, узлами-родителями детерминированных узлов могут выступать все типы узлов. В отличие от остальных видов узлов, детерминированный узел не стохастический и имеет уникальное значение как функция от значений узлов-родителей. Детерминированный узел может иметь три подтипа, в зависимости от выводов: категориальный, если выводит неупорядоченные значения, обозначенные комбинацией символов, дискретный если производит значения из интервала $[a,b]$, и непрерывный, если производит значения из множества вещественных чисел. В последнем случае непрерывного детерминированного узла задать значение узла (т.е. данные) невозможно, в иных случаях значение задается так же, как и в соответствующих типах стохастических узлов.

3. Процесс работы в Mbayes

В этом разделе мы рассмотрим процесс работы в MBayes, начиная с создания модели.

3.1. Загрузка и создание модели

Рассмотрим создание графа из файла данных в формате CSV, представленного в рис. 19. Видно, что этот файл содержит 3 колонки, одна из которых содержит значения [a1,a2,a3], другая содержит целые числа, а третья - вещественные. Для автоматического создания модели с узлами используем:

```

1 x,y,z
2 a2,3,2.596194881545987
3 a1,1,0.609941462477486
4 a1,0,-0.15459008207063918
5 a2,2,1.442516745205486
6 a2,4,2.741587422418413
7 a2,3,3.6371775663905823
8 a3,4,5.094089943277996
9 a2,1,2.0671617954915433
10 a2,2,3.0791701934751496
11 a2,3,3.0798647498261396
    
```

Рис. 19: Пример CSV файла с тремя переменными, одна в каждой колонке

СоздатьГраф → Загрузить из файла

После выбора файла и нажатия 'Определить узлы' модель будет содержать три несвязанных узла, представленных на рис 20, а распределения узлов будут иметь значения заданные по умолчанию, показанные в рис. 21, 22, 23.

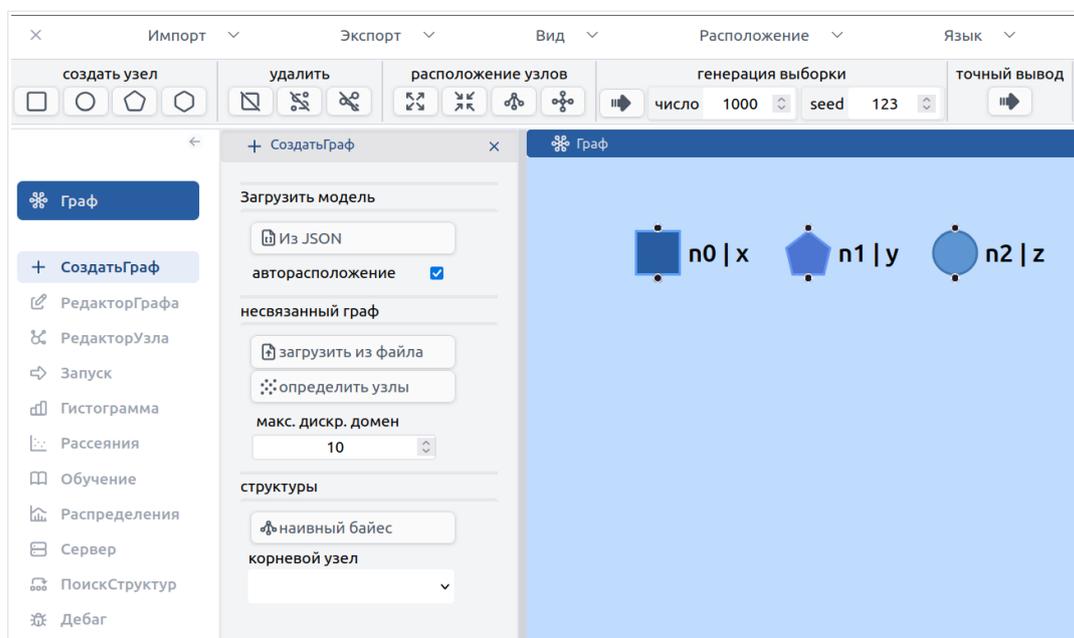


Рис. 20: Вид модели, полученной при автоматическом определении узлов из CSV файла

node: id: n0 type: categorical

node probability table

a1	-	a2	-	a3	-	+
0.33		0.33		0.34		$\Sigma=1$

Рис. 21: Значения, заданные по умолчанию для категориального узла 'x', созданного из данных

node: id: n1 type: discrete

node probability table

mu	distribution	lower_bound	upper_bound
1	poisson	0	4

Рис. 22: Значения, заданные по умолчанию для дискретного узла 'y', созданного из данных

node: id: n2 type: continuous

node probability table

distribution	loc	scale	lower_bound	upper_bound
normal	0	1		

Рис. 23: Значения, заданные по умолчанию для непрерывного узла 'z', созданного из данных

3.2. Редактирование модели

Далее рассмотрим, как можно редактировать модель, добавляя и удаляя узлы и связи, а также обучая параметры.

3.2.1. Создать узел

Для создания нового непрерывного узла используем

РедакторГрафа → **Непрерывный**

что создаст новый узел в модели. Удаление узла можно сделать выбрав узел, который требуется удалить и нажать 'Delete'/'Удалить' на клавиатуре или в верхней панели (раздел 2.1.1).

3.2.2. Добавить связь

Для того чтобы создать связь между узлами 'x' (n0) и 'z' (n2) требуется левой кнопкой мыши захватить нижний компонент узла n0, '●' (обозначим **красным**) и протянуть до верхнего компонента узла n2, '●' (обозначим **оранжевым**). Для удаления связи можно нажать на ○ в центре связи (обозначен **фиолетовым**), или выделить связь и нажать на 'Delete'/'Удалить' на клавиатуре либо в верхней панели (раздел 2.1.1).

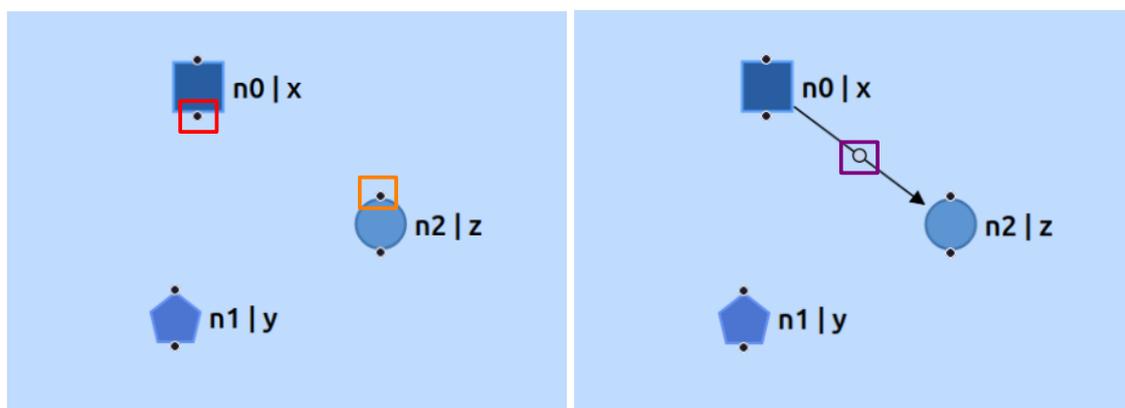


Рис. 24: Создание связи между узлами n0 и n2

3.2.3. Задать или обучить распределение

Задать форму условной зависимости в случае непрерывных и дискретных узлов можно через редактирование каждого из параметров условных распределений, основываясь на теоретических аспектах решаемой задачи или экспертных оценках. Также возможен процесс калибровки параметров условных распределений методом максимального правдоподобия с использованием:

Обучение → **Загрузить данные** → **Выучить параметры**

После обучения параметров модели таблицы условных вероятностей узлов в данном примере принимают значения, показанные на рис. 26 27 28

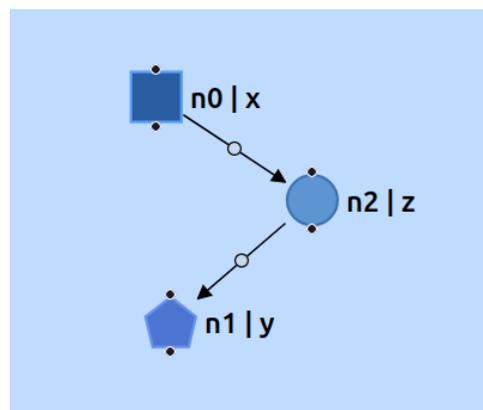


Рис. 25: Вид связей графа, на котором запускается процесс обучения параметров

узел: <input type="text" value="x"/>	id: n0	тип: категориальный				
Таблица условной вероятности						
<input type="text" value="a1"/>	-	<input type="text" value="a2"/>	-	<input type="text" value="a3"/>	-	<input type="text" value="+"/>
0.287138584; 0.471585244; 0.241276171						$\Sigma=1$

Рис. 26: Значения, полученные после процесса обучения для категориального узла 'x', созданного на данных

узел: <input type="text" value="y"/>	id: n1	тип: дискретный	
Таблица условной вероятности			
<input type="text" value="mu"/>	distribution	lower_bound	upper_bound
<input type="text" value="exp(0.0760)"/>	poisson	<input type="text" value="0"/>	<input type="text" value="4"/>

Рис. 27: Значения, полученные после процесса обучения для дискретного узла 'y', созданного на данных, где $\mu = \exp(0.0760 + 0.188 * n2)$

узел: z	id: n2	тип: непрерывный			
Таблица условной вероятности					
x	distribution	loc	scale	lower_bound	upper_bound
a1	normal	0.0681647	1.0262841		
a2	normal	3.0045470	1.0566261		
a3	normal	5.8860892	1.0771670		

Рис. 28: Значения полученные после процесса обучения для непрерывного узла 'z', созданного на данных

3.2.4. Ввести данные

Ввод 'Значения' в один или несколько узлов модели можно сделать из панели **РедакторУзла**. Пример для дискретного узла показан на рис. 29 где значение узла $n1=4$. Приложение автоматически добавляет значение поля в название узла на графе и изменяет цвет контура узла на **оранжевый** с целью его выделения в графе. Чтобы очистить значение узла требуется в панели **РедакторУзла** нажать .

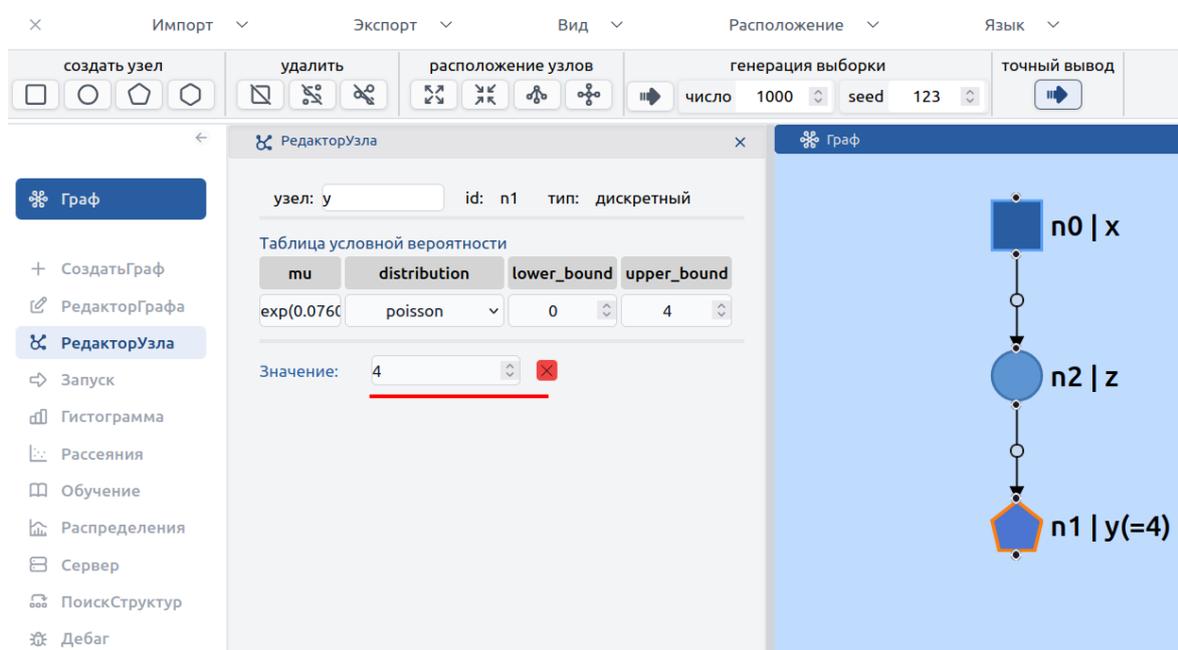


Рис. 29: Вид модели после ввода 'значения' в узел $n1$

3.3. Запуск вычислений

Запуск вычислений можно произвести используя:

Запуск → Генерация выборки

Это запустит алгоритм Монте-Карло с выбранным в панели 'Запуск' числом симуляций. Для повышения точности оценок можно повысить параметр 'число'. Рекомендованное максимальное значение размера выборки, при работе в приложении, 100,000. Если требуется более высокое количество симуляций, например при оценке редких событий, может потребоваться прямая интеграция с серверным приложением или пересмотр структуры или параметров модели. Параметр 'seed' контролирует инициализирующее значение ГСЧ используемое MBayes и повторяемость симуляций. Если требуется создать большую выборку то требуется изменять значение 'seed' для каждого запроса.

При успешном запросе и получении результат от сервера в приложении будет отображен статус запроса (рис. 30).

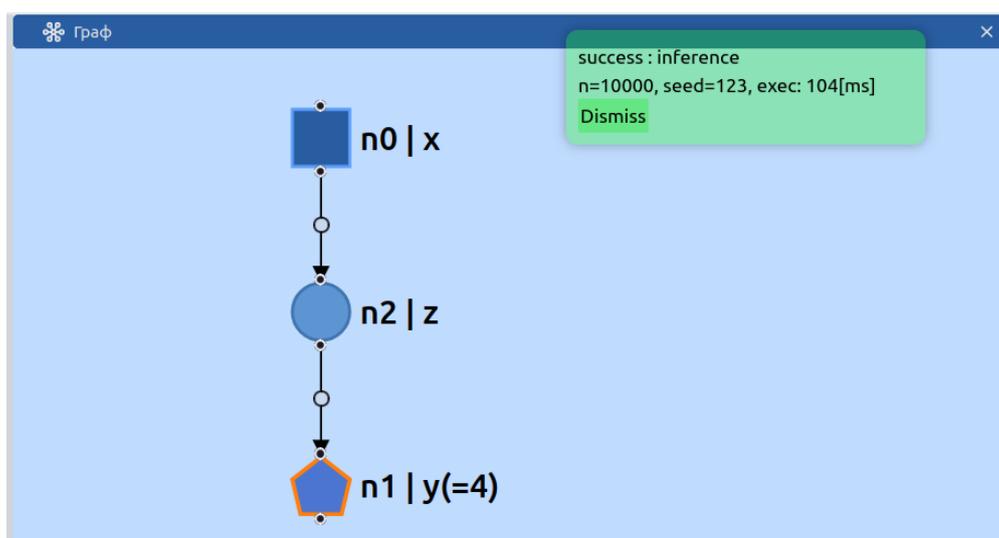


Рис. 30: Вид статуса после удачного запуска симуляции

При ошибке приложение отобразит статус запроса (рис. 31) и краткое описание причины.

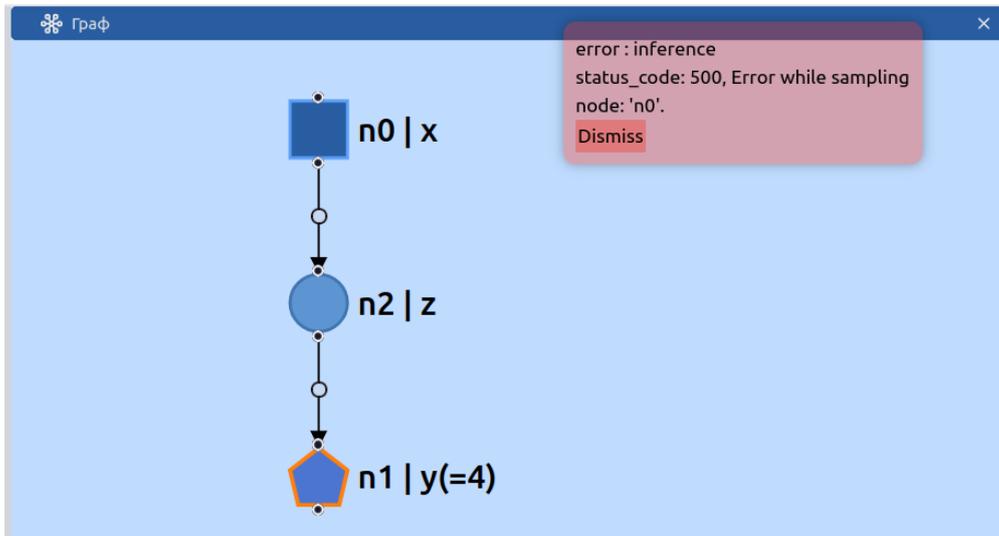


Рис. 31: Вид статуса после ошибки при запуске симуляции

3.4. Анализ результатов

При успешном запуске вычислений анализ результатов можно произвести в панелях

- Запуск
 - Эффективного размера выборки (Effective Sample Size/ESS)
 - Оценка маргинального правдоподобия (Marginal Likelihood Estimate)
- Гистограмма
 - Вывод статистик одномерного распределения
- Рассеяния
 - Вывод статистик двумерного распределения
- Распределения

В этом примере рассмотрим анализ априорного и апостериорного распределения узла ' $n_2 | z$ ', используя:

Распределения → **Добавить гистограмму** → **Загрузить данные из узла**

Получим отображение априорного распределения узла ' $n_2 | z$ ' на рис. 32.

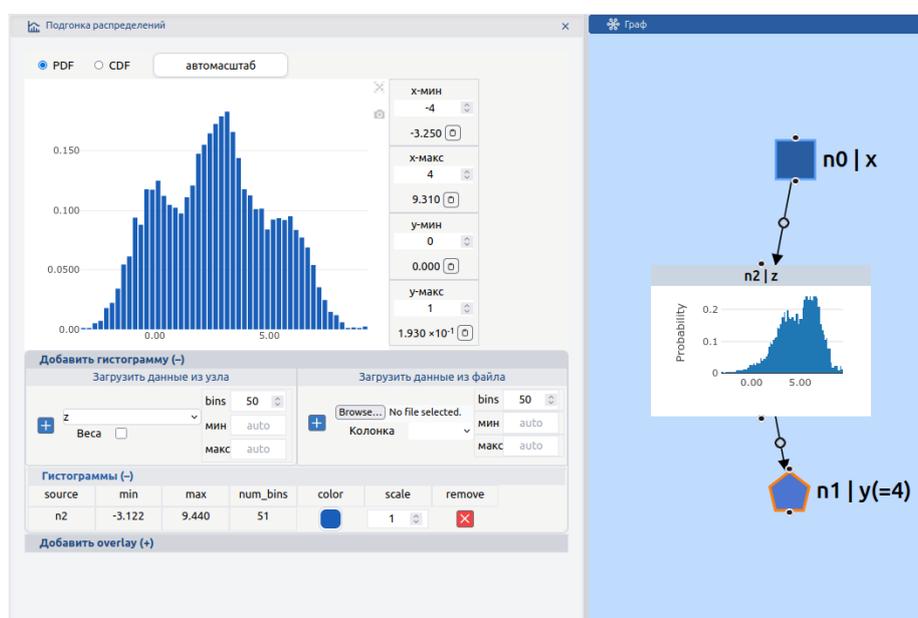


Рис. 32: Априорное распределение узла n_2

Далее добавляем три кривые плотности (overlay), используя:

Распределения → **Добавить overlay**

Получаем результат на рис. 33.



Рис. 33: Априорное распределение узла n2 с Гауссовскими кривыми плотности

Добавив гистограмму n2 с опцией 'Весы', можно получить оценку апостериорной вероятности. Также, используя overlay, можно получить оценку апостериорной плотности, как отображено в рис. 34.

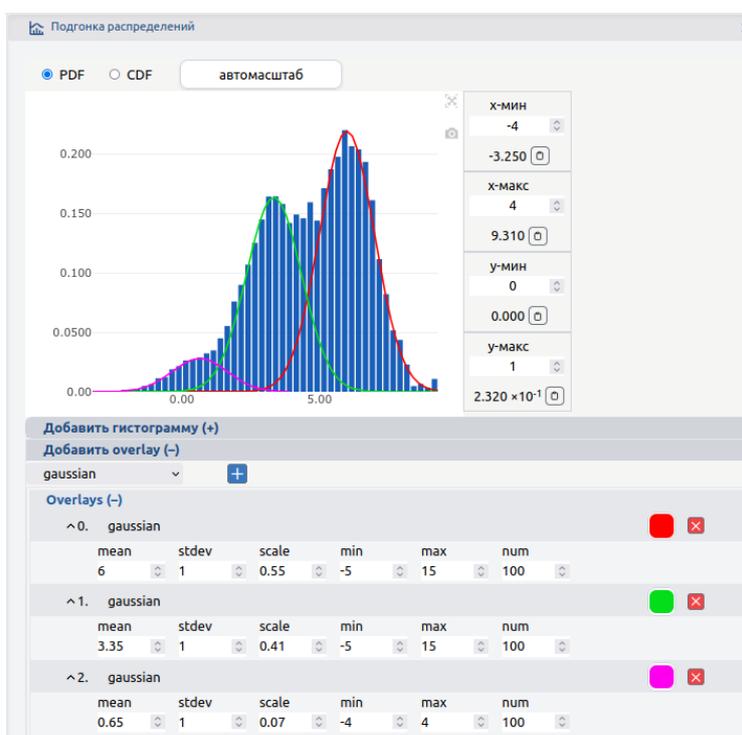


Рис. 34: Апостериорное распределение узла n2 с Гауссовскими кривыми плотности

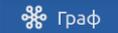
3.5. Экспорт модели

Для экспорта модели в формате JSON требуется использовать верхнюю панель редактора (раздел 2.1.1)

Экспорт → Модель или **Модель и метаданные**

Для экспорта значений симуляций в формате CSV файла, с колонками обозначающими узлы, требуется использовать:

Экспорт → Сэмпл

Для экспорта представления графа из раздела  **Граф** в формате PNG требуется использовать:

Экспорт → png image

Для экспорта гистограмм узлов (раздел 3.1.5) в формате JSON

Гистограмма → Скачать гистограмму

Также возможно экспортировать графики в формате PNG. Для этого следует нажать на символ  который отображен на каждом их графиков в правом верхнем углу.

3.6. Завершение работы

Для завершения работы требуется закрыть браузер в котором открыто приложение.

4. Синтаксис формул и детерминированных узлов

В таблицах условных вероятностей непрерывных и дискретных узлов поддерживаются следующие операторы и математические функции:

арифметические: +, -, *, /

степень: **

модуль: %

тригонометрические: sin, cos

математические: exp, sqrt, abs, log, log10

Функции могут быть использованы для построение функций условной зависимости узла от узлов-родителей.

Детерминированные узлы поддерживают следующий синтаксис построения функций (синтаксис в форме Бэкуса-Наура, BNF)

expr : SIGNED_NUMBER

- | CNAME
- | 'CNAME'
- | -CNAME
- | **expr** + **expr**
- | **expr** - **expr**
- | **expr** * **expr**
- | **expr** / **expr**
- | **expr** ** **expr**
- | **expr** == **expr**
- | **expr** != **expr**
- | **expr** > **expr**
- | **expr** >= **expr**
- | **expr** < **expr**
- | **expr** <= **expr**
- | **expr** && **expr**
- | **expr** || **expr**
- | ! **expr**
- | (**expr**)

compound_expr : **expr**

- | if(**expr**){**expr**}(elif(**expr**){**expr**})* else{**expr**}
- | max(**expr**,**expr** (,**expr**)*)
- | min(**expr**,**expr** (,**expr**)*)
- | match **expr** (case **expr** : **expr**)* default : **expr**
- | log(**expr**)
- | exp(**expr**)
- | abs(**expr**)

5. Установка, запуск и удаление приложения

Поскольку приложение является веб-сервисом, доступным через локальную сеть, установки на стороне пользователя не требуется. Доступ к приложению осуществляется по IP-адресу и порту, который конфигурирует администратор сети где установлено приложение.

Для запуска приложения пользователем требуется, в совместимом браузере, открыть ссылку:

```
http://<ip_addr>:<port>
```

где <ip_addr> - IP-адрес, а <port> - порт приложения.

Установка приложения в контур заказчика производится при поддержке специалистов MSистемы, порядок действий и процесс установки зависит конфигурации сети и инфраструктуры заказчика на котором будет работать приложение и обсуждается при покупке лицензии на ПО.

6. Решение типовых проблем (FAQ)

1. Не видно гистограмм/диаграмм рассеяния при смене узла

Повторно запустите вычисления в панели Запуск.

2. Не видно диагностик Эффективного размера выборки (ESS) и оценки маргинального правдоподобия

Убедитесь, что хотя бы одном узле модели задано Значение узла и повторно запустите симуляцию в панели Запуск

3. Гистограммы априорных и апостериорных распределений не отличаются.

Проверьте, что в модель введены значения, а также форму зависимостей между узлами с данными и узлов, где появляется такая ошибка — это может быть свойство созданной модели.

4. Ошибка при запуске инференса для узла, Error while sampling node 'm'.

Проверьте таблицу условной вероятности узла 'm', форму заданного распределения. Если узел категориальный, проверьте, что вероятности в каждой строке суммируются к единице и неотрицательны. Если узел непрерывный, проверьте, что параметры масштаба распределений не равны нулю.

5. Не могу ввести значения в детерминированный узел.

Проверьте, что подтип узла категориальный или детерминированный, в непрерывные детерминированные узлы ввод значений невозможен.

6. Ошибка при запуске точного вывода.

Проверьте, что граф содержит только категориальные узлы. Проверьте, что таблицы условных вероятностей не содержат невероятных значений (одно из заданных значений в узлах не имеет априорную вероятность, равную нулю) и нет конфликтов информации (комбинация значений в узлах имеет априорную вероятность, равную нулю).

7. Контакты и техническая поддержка

Оставить отзыв, пожелания или связаться по вопросам техподдержки:

Е-mail: m.systemi@yandex.ru

Телефон: +7 905 274 95 30,

Телеграм: @mbayes_ru

График работы: Пн-Пт, 10:00-17:00 МСК