

# НЕОПРЕДЕЛЕННОСТЬ В ОЦЕНИВАНИИ ПАРАМЕТРОВ РИСКОВАННОГО ПОВЕДЕНИЯ ИНДИВИДОВ

---

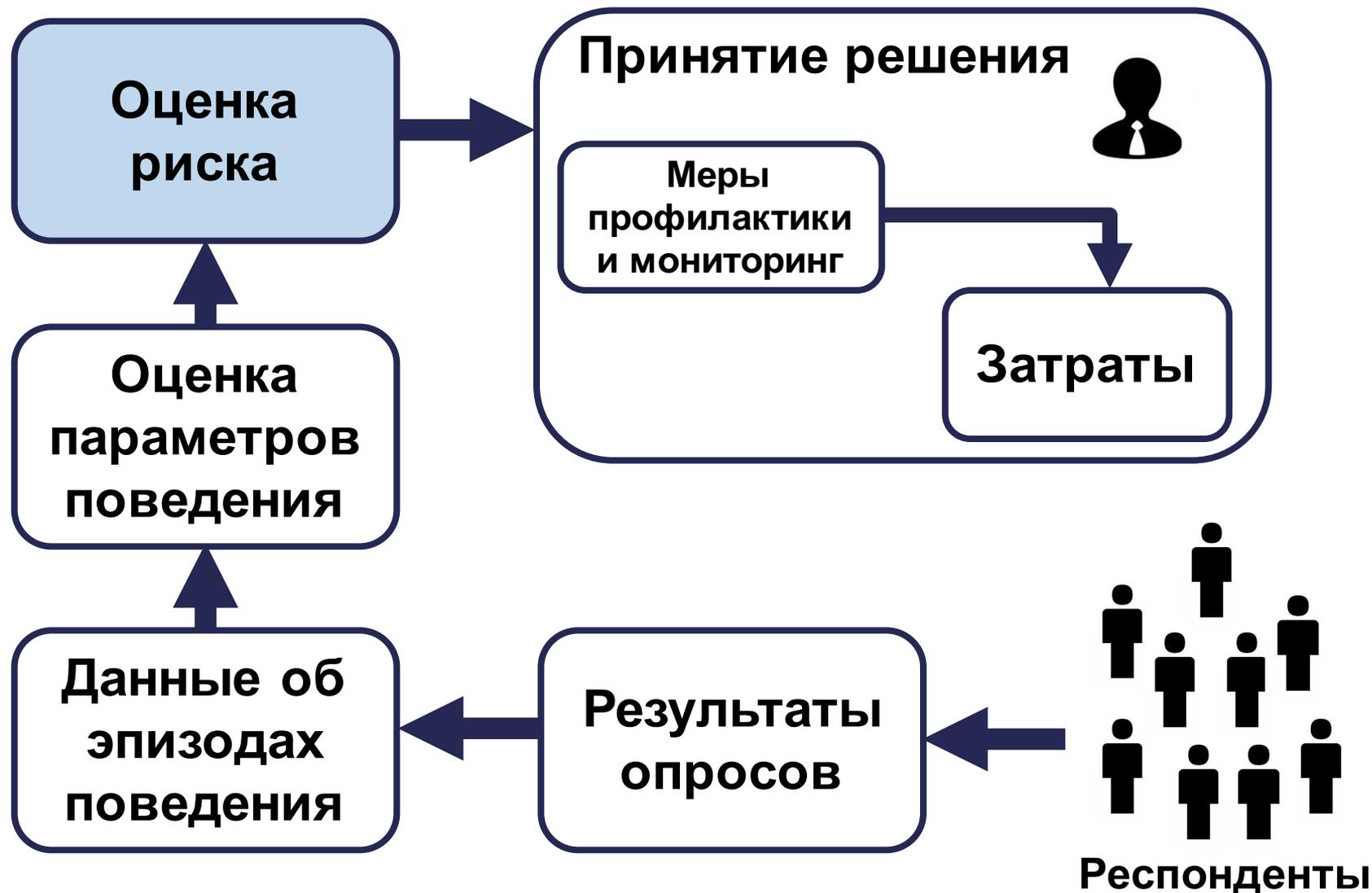
**Алёна Владимировна СУВорова**

к.ф.-м.н., СПИИРАН, [suvalv@gmail.com](mailto:suvalv@gmail.com)

**Александр Львович тулупьев**

д.ф.-м.н., СПИИРАН, [alt@iias.spb.su](mailto:alt@iias.spb.su)

# Оценка параметров поведения

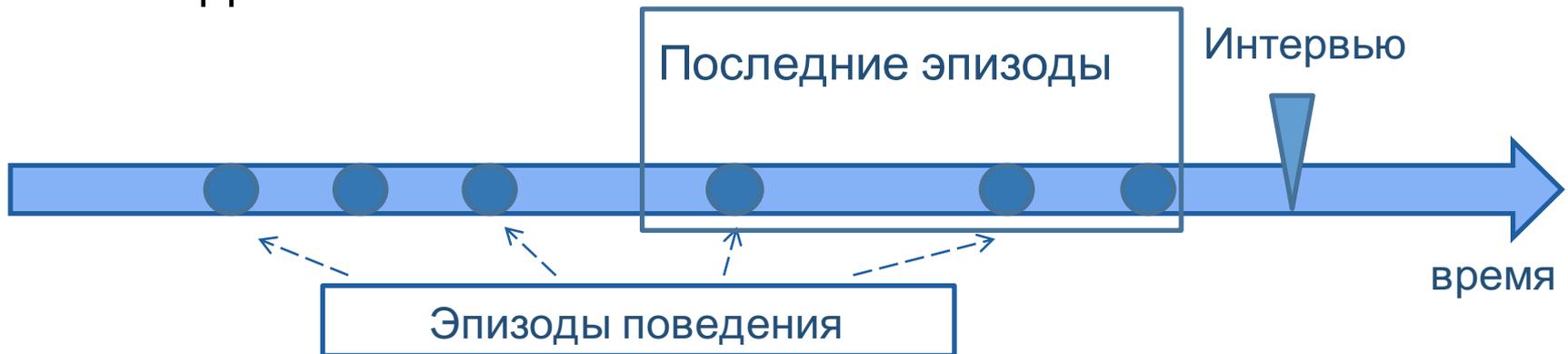


# Описание модели

- косвенное оценивание интенсивности поведения по данным о нескольких его последних эпизодах
  - дискретизация значений интенсивности и длин интервалов между эпизодами
- модель на основе байесовской сети доверия
  - сведение задачи к описанию более простых зависимостей
  - использование как статистических данных, так и экспертных сведений
  - использование существующего алгоритмического аппарата и свободно распространяемого программного инструментария

# Исходные данные

- Длины интервалов между последними эпизодами поведения



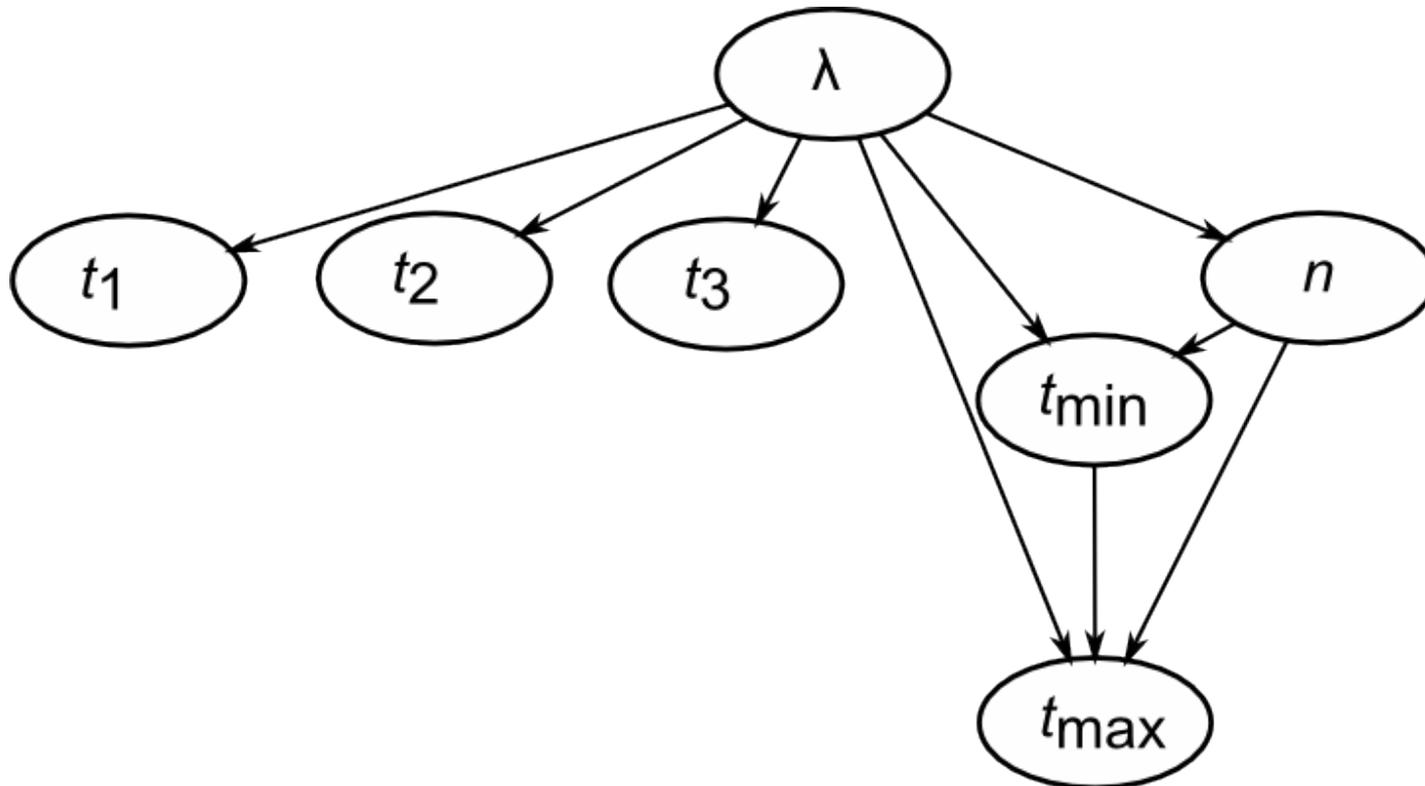
- + максимальный интервал между эпизодами
- + минимальный интервал между эпизодами

# Применение модели

Два типа предсказаний:

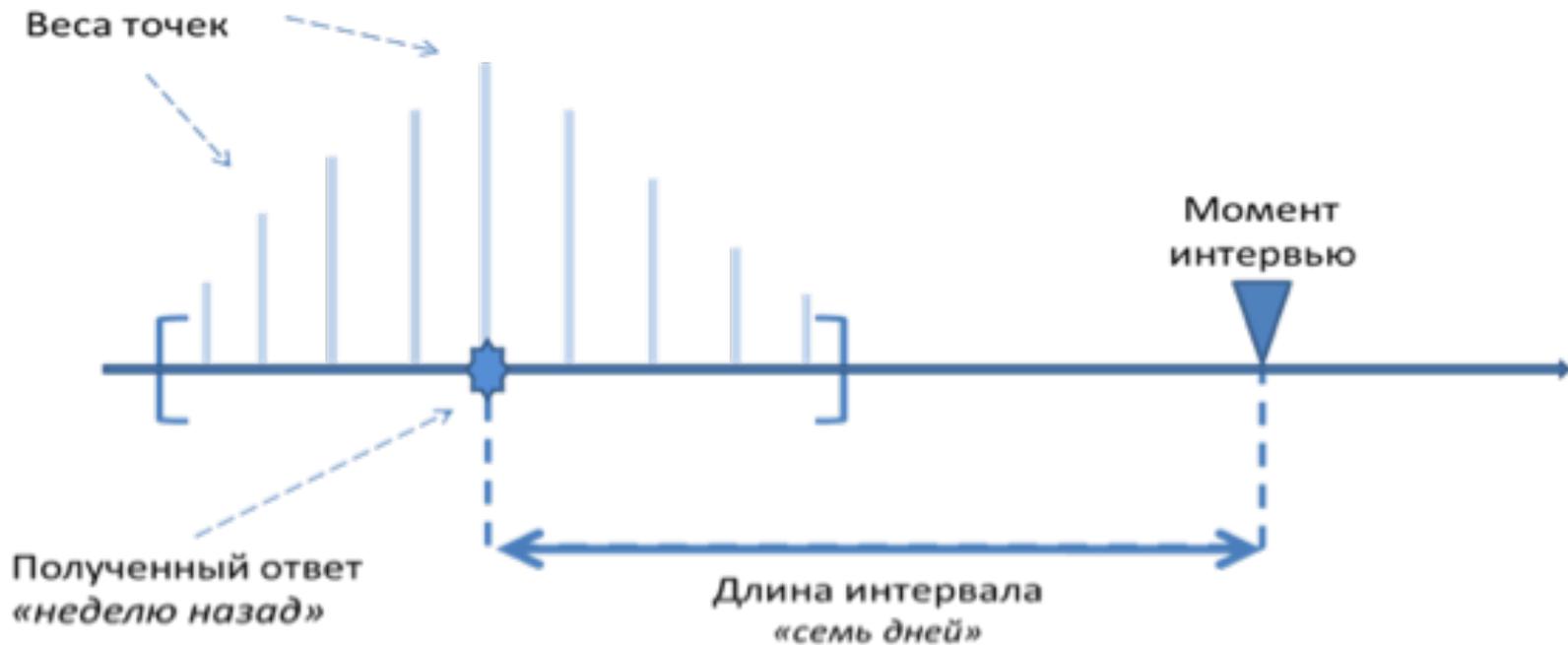
- Индивидуальные оценки интенсивности поведения
  - => персональные рекомендации
- Оценивание интенсивности поведения в группе
  - => планирование затрат, разработка политик, оценивание превентивных мероприятий

# Структура БСД: экспертно



# Неопределенность ответов

- Данные на естественном языке
- Сложности при воспоминании событий
- Не одно значение, а диапазон значений



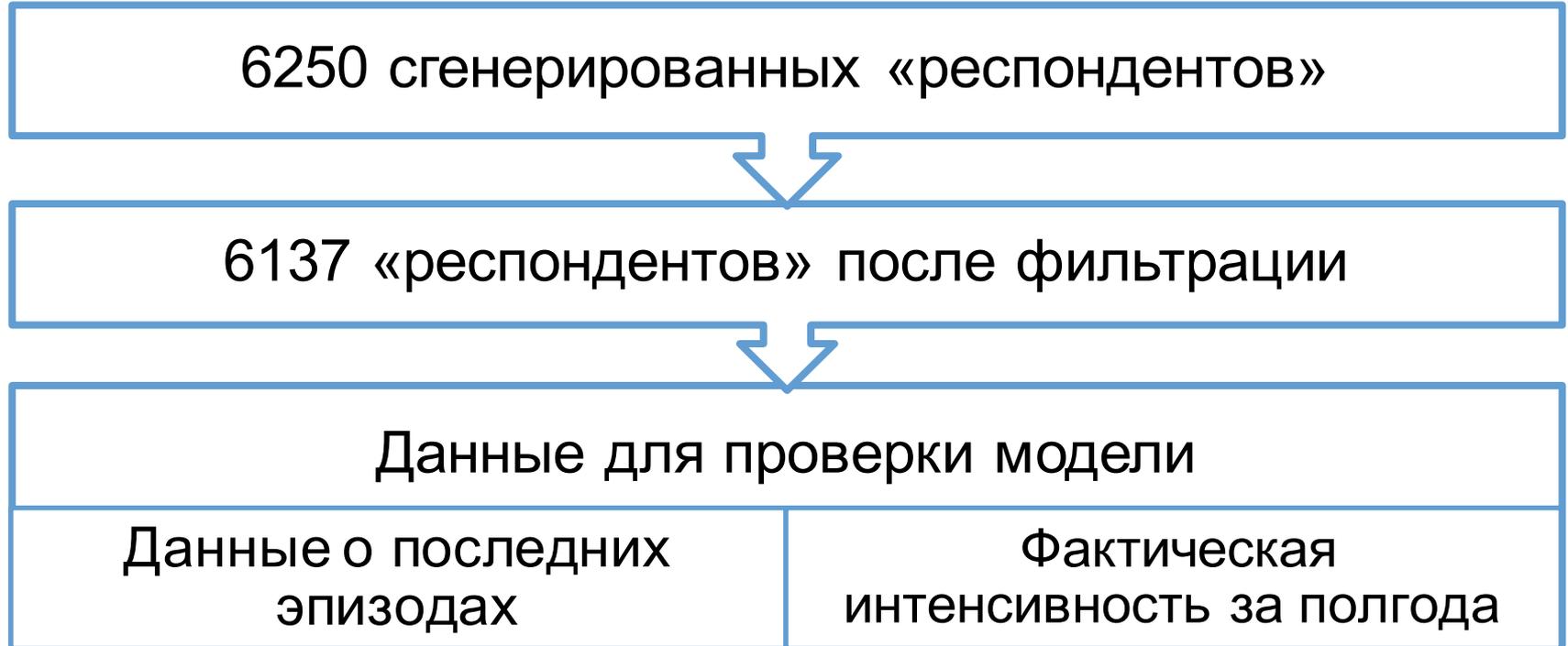
# Численные эксперименты

- Искусственно сгенерированные данные (в соответствии с теоретическими предположениями)
- Данные с шумом для моделирования неточного ответа
- Инструменты: R, bnlearn, gRain
- Вероятностные распределения в БСД построены на основе «чистых» данных
- Сравнение трех типов предсказаний
  - на «чистых» данных
  - на данных с шумом
  - на данных с шумом с коррекцией метода вычислений

# Численные эксперименты

- Рассмотрены только две переменные для предсказания
  - дискретные интервалы
  - интервальное свидетельство
  - несколько предсказаний для возможных пересечений
  - итоговое предсказание – согласно взвешенной сумме распределений

# Данные



Обучающая выборка: 5958 «респондентов» для построения условных вероятностей при определении модели

# Сравнение моделей: метрики

Алгоритм	Исходные данные		
	без шума	с шумом	
		без обработки	с коррекцией
Точность (accuracy)	0,531	0,499	0,514
Средняя точность (average accuracy)	0,883	0,874	0,878
Точность (macro precision)	0,466	0,431	0,449
Полнота (macro recall)	0,462	0,438	0,445

# Выводы

- рандомизация ответов позволяет учитывать возможную неточность исходных данных
  - без использования предположений о типе шума
  - повышение качества даже для двух переменных
- вычислительная сложность при большом числе переменных
- возможное решение: использование виртуальных свидетельств

# НЕОПРЕДЕЛЕННОСТЬ В ОЦЕНИВАНИИ ПАРАМЕТРОВ РИСКОВАННОГО ПОВЕДЕНИЯ ИНДИВИДОВ

---

**Алёна Владимировна СУВорова**

к.ф.-м.н., СПИИРАН, [suvalv@gmail.com](mailto:suvalv@gmail.com)

**Александр Львович тулупьев**

д.ф.-м.н., СПИИРАН, [alt@iias.spb.su](mailto:alt@iias.spb.su)