# Моделирование временных рядов
# Лекция 1
# Введение

Борис Демешев, Матвей Зехов

Временные ряды по сути своей лишь частный случай стандартной задачи регрессии или классификации. Они возникают в большом количестве областей. Например, любая компания из ретейла будет прогнозировать количество товаров, которые необходимо поставить в магазин или даркстор. Также обычно сразу приходит на ум котировки на фондовой бирже. Из физики и инженерных задач возникает анализ сигналов от датчиков. Введём определение временного ряда. Перед этим мы должны ввести (очень нестрого) понятие случайного процесса.

## 1 Определения

Случайный процесс – это некоторая *последовательность* случайных величин $Y_t$. Мы специально сконцентрируемся только на дискретных последовательностях с вещественными значениями, так как для большинства задач этого достаточно. Следовательно, временным рядом мы будем называть некоторую *реализацию* случайного процесса, $y_t$. Иногда эту реализацию ещё называют траекторией. Также можно встретить определение, что временной ряд это и есть случайный процесс (последовательность случайных велечин), но с практической точки зрения это немного не интуитивно и мы постараемся этого избегать.

На практике обычно мы имеем дело с последовательностями с конечным числом элементов $(y_t)_{t=1}^{T}$, где T – количество наблюдений. Иногда в учебных целях иногда будем затрагивать последовательности с бесконечным числом элементов: $(y_t)_{t=1}^{t=+\infty}$ или $(y_t)_{t=-\infty}^{t=+\infty}$

В классических моделях машинного обучения мы предполагали наблюдения в обучающей выборке независимыми и одинаково распределёнными: $X = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$. Однако от этих предпосылок нам придётся отказаться. Почти всегда элементы последовательности будут зависимы между собой и нашей задачей будет выяснить характер этой связи. Грубо говоря, нам необходимо восстановить характеристики случайного процесса по сгенерированной тракетории. Да, наличие структуры в данных не позволяет нам напрямую использовать стандартные техники машинного обучения, но в то же время из этой структуры можно выделить много дополнительной и полезной для прогнозирования информации.

Можем ли мы в принципе быть уверены, что сможем его восстановить? В какой-то мере ответ на это даёт Теорема Дуба о разложении [1]. Говоря в очень упрощённых терминах, она говорит о том, что почти любой "хороший" процесс можно разложить на прогнозируемую (детерминированную) и принципиально непрогнозируемую части. Следственно, мы никогда не сможем восстановить процесс идеально. Но тем не менее, часть процессов более склонна к детерминированному поведению, а часть – менее. Например дневная температура яввляется очень сезонной величиной, то есть имеет паттерн, удобный для прогнозирования. Мы будем учиться обнаруживать и выделять такие паттерны в данных. С другой стороны, котировки акций наиболее близки к хаотичному движению и весьма трудно поддаются прогнозированию. Этим занимаются скорее в области количественных финансов. В нашем курсе котировки акций могут быть рассмотрены скорее из-за удобства и качества данных, но не более чем для иллюстрации. Исключение составит тема прогнозирования волатильности.

# 2 Возможные постановки задач

Задачи на временных рядах можно рассматривать с двух сторон. Во-первых, их можно свести к стандартным методам машинного обучения с минимальными оговорками в подготовке данных. Во-вторых, можно рассматривать это направление как развивавшееся независимо в контексте эконометрических задач с уклоном в логику описания данных. Мы кратко поговорим про первый подход и более подробно про второй.

## §2.1 Случай одного ряда

## §2.2 Случай набора рядов

# Список литературы

[1] https://wikichi.ru/wiki/Doob_decomposition_theorem