

Предисловие

Перед вами курс лекций Альберта Николаевича Ширяева, излагающий одну из самых замечательных теорий анализа данных — теорию, объясняющую, как по наблюдениям за реализацией случайного процесса обнаруживать «спонтанно возникающие эффекты», когда свойства процесса изменяются скачком. Большая часть проблем анализа данных связана с исследованием стохастических динамических систем, и, конечно, обнаружение в них существенных, но редких событий часто имеет решающее значение. Это верно и для анализа финансовых потоков, и для анализа данных с датчиков сейсмических станций предварительного оповещения о возможных землетрясениях, и для мониторинга динамических факторов, характеризующих изменения качества поисковой Интернет-системы, и для слежения за сигналами жизнедеятельности больного в палате интенсивной терапии, и для многих других приложений, где редкие скрытые события могут принципиально изменить законы поведения рассматриваемой сложной системы. Оказывается, можно обнаруживать и даже предугадывать такие редкие события, и курс А. Н. Ширяева является базовым для того, кто хочет научиться это делать.

Два слова из истории. А. Н. Ширяев излагает только одну из известных теорий анализа стохастических динамических данных. Другие теории, как, например, теория скрытых марковских моделей или теория обучающихся марковских сетей, также широко применяются для обнаружения редких событий. Однако среди всех этих теорий излагаемая в настоящем курсе является наиболее теоретически обоснованной, наиболее конструктивной и наиболее востребованной на практике. Ее называли раньше теорией обнаружения разладок в сложных системах. В настоящее время ее включают в теорию оптимальных правил остановки, дающую возможность строить процедуры обнаружения моментов изменения свойств про-

цесса с минимальной задержкой и минимальным средним числом ложных тревог.

Эта теория во многом создана автором курса, А. Н. Ширяевым, так что читатель получает здесь ее изложение из первых рук. Отмечу сразу, что теория эта сложна даже для студентов вероятностных специализаций. На протяжении трех лет ее преподавания в Школе анализа данных Альберт Николаевич несколько раз менял изложение курса, стараясь сделать его максимально доступным для студентов инженерных специальностей. Настоящая книга — результат этих поисков. И вместе с тем не следует думать, что эту книгу будет легко изучать инженеру. Для тех, кто действительно захочет это сделать, можно дать несколько советов:

1) необходимо изучать параллельно с освоением материала этой книги учебники А. Н. Ширяева «Вероятность-1» и «Вероятность-2»,

2) следует иметь в виду, что книга написана очень плотно и поэтому разбирать ее надо медленно, но систематически,

3) необычную роль в книге играет раздел приложений — в него надо систематически заглядывать,

4) не нужно жалеть времени на первые 40 страниц, где дается «вспомогательный» материал, классические теории Неймана—Пирсона и Вальда,

5) не пропускайте лекции и семинары на курсах Яндексa и активно работайте на них, а главное — если вы понимаете, что эта теория вам нужна, будьте уверены в том, что вы ее освоите, хотя поначалу может казаться, что это не так.

Следует отметить исключительную продуманность и ясность изложения материала. А. Н. Ширяев уделил в книге много места случаю дискретного времени. Уверен, что для инженера это наиболее удобно и понятно, несмотря на то, что иногда вычисления, связанные с дискретностью, оказываются довольно сложными.

Еще одно замечание. Параллельно с освоением теории и методов обнаружения редких событий, читатель-инженер узнает из этой небольшой книжки о двух фундаментальных моделях случайных процессов — о броуновском движении и марковских процессах. Тот, кто освоит их, получит твердую базу для своих собственных исследований в самых различных приложениях.

В заключение хочу обратиться к тем, кто смотрит на теорию анализа данных как на теорию, доставляющую средства, которые помогают в деле разработки новых Интернет-технологий. В настоящее время практически все наши исследования базируются на статическом статистическом анализе. Очень трудно, если вообще возможно, найти исследование процессов в Интернете, которое базировалось бы на стохастическом аппарате теории случайных процессов. Между тем, именно динамика случайности в наших системах представляет наибольший интерес. Можно надеяться, что будущие разработчики Яндекса, изучившие курс А. Н. Ширяева, найдут адекватные методы исследования поиска и других процессов, которые мы разрабатываем.

И. Б. Мучник