Автор:
Петухов Виктор, гр. 3512

Количественное описание моделей

§1. Понятия измерения

Измерение — алоритмическая операция, которая конкретному наблюдаемому состоянию объекта ставит в соответствие определенное обозначение (число-номер-символ).
Существуют различные подходы к измерению.

За последние годы в поле зрения науки попали наблюдаемые явления в принципе не допускающие числовой меры (слабые или качественные шкалы).
Не четкие наблюдения рассматриваются не как ошибка измерения, а как неотьемлимое природное свойство.

§2. Теоретико-множественные отношения - базис количественного описания моделей

Наиболее важными для теории измерений являются следующие отношения:
- Эквивалентность
- Порядок
- Толерантность

Все они выражаются через комбинацию базисных свойств: рефлекивность, транзитивность, симметричность.

Отношение R называется **рефлексивным**, если выполняется $\forall$ x $\in$X xRx (наприер, x=x и т. п.)
Отношение называется **транзитивным**, если {a*b, b*c} => a*c
Отношение S называется **симметричным**, если xSy => ySx
Отношение **антисимметричности**: если {xSy, ySx} => x=y

Эквивалентность — это значит рефлексивность, симметричность и тразитивность.

Отношение порядка бывает двух типов:
- Транзитивность и антисимметричность — **отношение строгого порядка**
- Транзитивность, антисимметричность и рефлексивность — **отношение не строгого порядка**
- Транзитивность, не рефлексивность и антисимметричность — **отношение толерантности** (любые два подмножества имеют какие-либо общие признаки).

Классические измерения предполагают, что на множестве состояний наблюдаемого объекта, а также на множестве наблюдений может быть построено отношение эквивалентности (все сосотояния, над которыми производятся наблюдения, должно быть различимы).

§3. Шкала наименований (номильная шкала)

Считаем, что число различимых состояний объекта конечно (перечислимо).

Каждому классу эквивалентности ставим в соответствие свое обозначение, тогда измерение: провести эксперимент над объектов, определить принадлежность результата к конкретному классу эквивалетности, записать результат с помощью символа, обозначающего класс. Указанное множество символов образует шкалу.

Обозначение классов: слвова естественного языка, произвольные символы (гербы, флаги, эмблемы), числа (если они имеют смысл номеров).

При большом или не фиксированном числе классов можно ввести иерархичность в обозначениях.

Номинальные шкалы на непрерывных множествах

На непрерывных множествах можно ввести номинальные шкалы искуственным путем. При работе с такими измерениями это нужно осознавать. Пример: номер цвета.

Порядковые (ранговые) шкалы

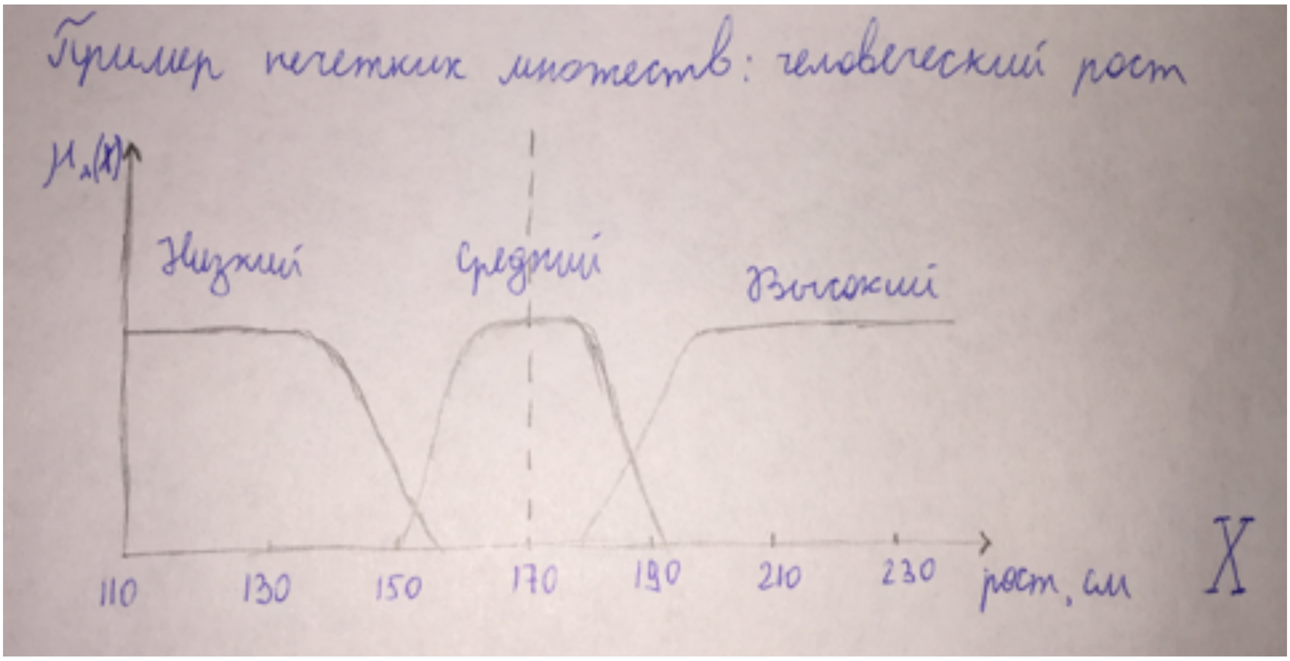
1) **Шкала простого порядка** удовлетворяет аксиомам тождества, а также аксиомам упорядоченности. A > B —> B < A A > B, B > C => A > C
2) **Шкала слабоого порядка**: A <= B, либо B <= A, A >= B, B >= C => A >= C. Иногда оказывается, что не каждую пару классов можно упорядочить по предпочтениям, а именно — некоторые пары считаются равными.
3) **Шкала частичного порядка** — имеются пары классов, не сравнимые между собой. Пример: изучение покупательского спроса.

Модифицированные порядковые шкалы

1) **Истиые порядковые шкалы** принципиально дискретны, тем не менее хочется применить порядковый принцип к принципиально непрерывных величинам (сила землетресения, твердость вещества и т. п.). Примеры: шкала твердости по Моосу (шкала устанавивает искуственный слабый порядок и границы твердости не имеют числового характера), шкала силы ветра по Бофорту, шкала магнитуд землетрясений, бальная оценка знаний учащихся (принципиально оценки по этой шкале справдливы только внутри этой генеральной совокупности).
2) **Интервальная шкала** — упорядочивание объектов можно выполнить настолько точно, что известны расстояния между любыми объектами (объективно-равные интервалы измеряются одинаковыми по длине отрезками шкалы), следовательно, введенные шкалы могут изменить произвольное начало отчета и произвольную единицу длины.
3) **Шкала отношений** — отношение двух наблюдаемых значений измеряемой величины не зависит от единицы измерения, принятой в шкале (то есть измеряемые величине в шкале отношений имеют естественный абсолютный ноль).
4) **Шкала разностей** (периодическая/циклическая шкала) — это шкала, значение которой не изменются при любом сдивиге на период.
5) **Абсолютная шкала** имеют абсолютный ноль и абсолютную единицу (основой абсолютных шкал является числовая ось — она используется как измеретильная ось в явном виде при пересчете предметов, а как вспомогательное средство присутствует во всех остальных шкалах).

Нечеткое описание ситуации

Нечеткое множество A состоит из неопределенного числа элементов X, причем по крайней мере некотоыре элементы из X могут как принадлежать множеству X, так и не принадлежат ему.



Ф-ия принадлежности $\mu_a(x)$ — число в интервале [0..1], выражающее степень отношения элемента к данном нечетному множеству. Если Ми принимает только либо 0, либо 1 — то множество четкое (тогда наше нечеткое множество А определяется парой числе: x и $\mu_a(x)$ — на этой двойке можно задать математические операции (пересечение, включение и т. п.)). Каждая измерительная шкала имеет свой нечеткий аналог. Однако для ряда объектов (красота, уровень знаний) вообще не существует объективной физической шкалы.

Функция принадлежности μ задается экспертным путем и не имеет связи со статистическими характеристиками объекта.
Задаваться она может следующими способами:

1) Организуется иерархия нечетких множеств (организуется такое нечеткое множетсво, у которого значение функции принадлежности также является нечетким множеством — тогда составная лингвистическая переменная определяется рекуррентно через лингвистические переменные низших порядок). Недостаток — мы теряем комонент оценки, связанный с взаимозависимостью признаков.
2) Организуется базовые конечные нечеткие множетсва, линейно-упорядоченные на произвольной оси (например, на оси бальных оценок). Упорядочение справедливо только для данного базового множества. Добавление к нему нового элемента требует переупорядичение всего множества.

Существует несколько подходов к определению функции μ:

1) Эвристический подход — субъект сам определяет, как он понимает степень принадлежности. Соотв., ф-ии μ разных экспертов будут отличаться.
2) Статистический подход — усредняем ф-ии μ, задаваемые рядом экспертов.
3) Частичное задание μ для нескольких значение x и последующее доопределение ф-ии подходящим методом. На графике: ставим отдельные точки, а все остальные участки соединяем линиями.
4) Интервальное задание (напр., задание пессимистиеской и оптимистической границы). На графике: ставим только границы.
5) Кратная нечеткость — задание μ, как нечеткого множества с помощью ф-ии принадлежности второго порядка.

§4. Границы применимости теории вероятностей в описании случайных событий

Существует несколько точек зрения на природу случайности, причем каждая из них имеет достаточное основание. Случайным нам представляет то, про что мы ещё пока не знаем закономерности.

1) **Случайность** — объективное свойство всех явлений.
2) **Промежуточная точка зрения** — существуют детерминированные явления и существует в принципе случайные явления.
3) **Синергетика** — сулчайные и детерминированные процессы сменяют друг друга в истории любой системы.

Теория вероятностей имеет дело в только такими случайности, которые подчинются строгой закономерности, выражаемой функцией распределния вероятностей. многие эксперементальные ситуации могут быть хаотическими, но не вероятнсотными.

§5. Способы задания вероятностных характеристик случаных процессов

Носитель распределения — некоторый материальный процесс (X), порождающий множество значений (x), имеющий определенную вероятность появления.
Примеры:
- Телефонные системы (случайным процессом является речь — непрерывное изменение звукового давления на микрофоне)
- Телевизионные системы (изображение — яркость определенных пикселей)
- Финансовые системы (курс валют)

С точки зрения преобразования информации эти процессы можно объединить термином «сигнал».
Сигналы несут сообщения (являются информационно-наполненными). Сообщения могут быть функциями времени или нет.

Сообщения являются либо случайной величиной, либо случайной функцией. Детерминированные сообщения не содержат информации (точнее — прием обеспечивает 1 бит информации). Поэтому в качестве стандартной модели сообщения используется случайный процесс или процесс со случайными параметрами, а отдельное сообщение рассматривается как его реализация.

Случайный процесс в общем случае задается n-мерной плотностью распределения вероятности: $\omega_n(x_1, x_2, \dots, x_n)$ $n \to \infty$
Можно записать: $\omega_n(x_1|x_2|x_3,\dots)$
При задании функции ω_n и тем более при её эксперементальной оценке возникают огромные трудности точностного и ресурсного порядка.
Поэтому стремятся упростить модель, используя различные приближения и допущения.

Виды приближения:

1) Вычисления, эксперементальная оценка не полной ф-ии ω_n, а её моментов:
   1) Математическое ожидание случайного процесса $m_x(t) = M[X(t)] = \int x\omega(x, t)dx$ (from -∞ to ∞)
   2) Корреляционная функция КФ = $R_{x_1,x_2} = M[X(t_1), X(t_2)]$ — она определяет степень сходства данной реализации с её сдвинутой во времени копией. Следовательно,

КФ позволяет выделить в сигнале скрытые периодичности. Значение корреляционной функции при t_1 = 2 (при Δt = 0) имеет максимум и называется дисперсией случайного процесса. Возможно конструирование моментов более высокого порядка (3-го и т. д.). По большей части ограничиваются моментами 1-го и 2-го порядков.

(можно вставить картинку)

2) Стационализация процесса. Стационарность случайного процесса, значит, что его статистическое описание не зависит от начала отчета. Это преобразование координат: *(вставить картинки)*. Иначе говоря, стационаризация требует креативного назначения и обоснования функции нестационарного тренда. В этом случае стандартное преобразование предусмотрено в любом статистическом пакете.
3) Допущение эрбодичности процесса. Эрбодичность — знвчит, что характеристики процесса, найденные усреднением по ансамблю реализации и по одной бесконечно длинной реализации совпадают.
4) Использование в качестве функции распределения ω_n стандартных форм, имеющих аналитическое представление с подбором параметров по результатом эксперементальной оценки реального процесса. Самое часто используемое распределение — нормальное (Гауссовский процесс). Гаусовский случайный процесс исключительно привлекателен для моделирования тем, что он полностью определяется двумя первыми моментами распределения (мат. ожиданием и дисперсией). Говорить о нормальном распределении конкретного случайного процесса можно только при выполнении центральной предельной теоремы (нормальный процесс формируется под действием суммарного влияния большого числа независимых случайных величин, каждая из которых вносит маленький вклад). Предложен целый ряд других распределений, имеющий аналитическое представление, у которых существуют отличные от нуля высшие моменты. В качестве примера назовем распределение $\chi 2$ — распределение суммы квадратов нормальных случайных величин. Оно имеет много частных вариантов (распр-ние Максвелла, распределение с длинным хвостом и т. п.). Рассмотрим пример: заложение фундамента для дома *(вставить картинку)*. В зависимости от того, какую функцию распределения мы считаем требующей реализации (отработки) принимаются те или иные конструктивные решения.

§6. Источники ошибок при применение статистических методов

Причины неправльного применения статистических методов не многочислены и всех их нужно учитывать.

1) Статистический вывод по своей природе случаен и статистики стремятся, чтобы каждая её процедура сопровождалась оценкой её качества. Для получения строгих оценок качества измерения требуется знать распределение ошибок измерения. Распределение ошибок содержит моменты более высоких порядков, чем исходное и оценивается с ещё большими погрешностями. Пример: если независимые случайные величины подчиняются нормальному распределению, то распределение вероятности суммы квадратов этих величин — $\chi 2$ — распределение, имеющее три момента. Мы измеряем мат. ожидание, его ошибка распределена по нормальному закону, но мы измеряем дисперсию, а она распределена по $\chi 2$. Доверительный интервал — ошибка дисперсии. Следовательно, от моментов типа мат. ожидание и дисперсии мы все ближе уходим к оценке непосредственно функции распределения (многомерной ф-ии с n высшими моментами). Методы ужесточения к качеству статистического вывода:
    1) Увеличение объема выборки (много раз измерить одну и ту же величину, ф-ию или значение). Этот путь ограничен размерами имеющейся генеральной совокупности.

2) Увеличение размерности задачи (использовать большее количество показателей). Этот путь ограничен тем, что с ростом числа показателей они становятся взаимосвязанными (кореллированными) — матрица задачи становится плохо определенной. Матрица задачи — основной определитель системы уравнений, связывающий все переменные. Выход — оценки эвристического типа.

2) Качество решений на выходе статистической процедуры зависит от того, что подается на вход. Возможны осознанные искажения, а также добросовестные заблуждения, а также возможны неудачные первичные опросники. Основной путь противодействия — обработка результатов всей выборки по частям и сравнение идентичности полученных резултьатов.
3) Статистической выборке могут подвергаться данные вообще не имеющие статистической природы. Меры борьбы — грамотная постановка задачи (то есть нужно искать такую зависимость, которая по природе своей является статистической).
4) Использование неадекватной статситической процедуры, не соответствующей действительному уровню априориной информации. Практически любой эксперементальный результат можно апроксимировать любой желаемой кривой с потерей точности. Если имеющийся уровень априорной информации мал, то лучше не использовать сложные ф-ии распределения или вообще использовать не параметрическую статситику. Хорошо в этой ситуации работает метод наименьших квадратов.
5) Неверное содержательная интерпретация правильного статистического вывода. Реакция — искать надежную, лучше физическую, причину закономерности.
6) Пропущенные значения. Их необходимо восстанавливать. Существуют методки борьбы, которые позволяют это делают.
7) Зашумленные данные. Существуют методки борьбы, которые позволяют это делают.

Не смотря на все предосторожности, в статистической практике возможны принципиальные ошибки, поэтому желательны организационные страховочные меры. Это выражается в том, что от показателей надежности переходят к показателям рисков.