Введение

Теория вероятности есть математическая наука, изучающая закономерности в случайных явлениях. Что же понимается под случайными явлениями? При научном исследовании физических и технических задач, часто приходится встречаться с явлениями особого типа, которые принято называть случайными. Случайное явление - это такое явление, которое при неоднократном воспроизведении одного и того же опыта протекает несколько по-иному. Приведем пример случайного явления. Одно и то же тело несколько раз взвешивается на аналитических весах: результаты повторных взвешиваний несколько отличаются друг от друга. Эти различия обусловливаются влиянием различных второстепенных факторов, сопровождающих операцию взвешивания, таких как случайные вибрации аппаратуры, ошибки отсчета показаний прибора и т.д. Очевидно, что в природе нет ни одного физического явления, в котором не присутствовали бы в той или иной мере элементы случайности. Как бы точно и подробно ни были фиксированы условия опыта, невозможно достигнуть того, чтобы при повторении опыта результаты полностью и в точности совпадали. Случайности неизбежно сопутствуют любому закономерному явлению. Тем не менее, в ряде практических задач этими случайными элементами можно пренебречь, рассматривая вместо реального явления его упрощенную схему, т.е. модель, и предполагая, что в данных условиях опыта явление протекает вполне определенным образом. При этом из бесчисленного множества факторов, влияющих на данное явление, выделяют самые главные, основные, решающие. Влиянием остальных, второстепенных факторов просто пренебрегают. Изучая закономерности в рамках некоторой теории, основные факторы, влияющие на то или иное явление, входят в понятия или определения, которыми оперирует рассматриваемая теория. Как и всякая наука, развивающая общую теорию какого-либо круга явлений, теория вероятностей также содержит ряд основных понятий, на которых она базируется. Естественно, что не все основные понятия могут быть строго определены, так как определить понятие - это значит свести его к другим, более известным.
Помимо понятия события и вероятности, одним из основных понятий теории вероятностей является понятие случайной величины. Случайной величиной называется величина, которая в результате опыта может принять то или иное значение, причем неизвестно заранее какое именно. Случайные величины, принимающие только отдельные друг от друга значения, которые можно заранее перечислить, называются прерывными или дискретными случайными величинами. Случайные величины, возможные значения которых непрерывно заполняют некоторый промежуток, называют непрерывными случайными величинами. Например, ошибка взвешивания на аналитических весах. Отметим, что современная теория вероятности преимущественно оперирует случайными величинами, а не событиями, на которые в основном опиралась "классическая" теория вероятностей. Корреляционные моменты, коэффициент корреляции - это числовые характеристики, тесно связанные во введенным выше понятием случайной величины, а точнее с системой случайных величин.

В моей курсовой работе рассмотрены такие важные понятия, как корреляционное отношение и индекс корреляции. Сама корреляция будет рассмотрена на примере. Для рассмотрения своих вопросов и вычисления индекса корреляции использовал формулы уравнения регрессии, вычисления дисперсии, корреляционной зависимости и детерминации.

Глава 1. Корреляционное отношение и индекс корреляции

1.1 Основные понятия

Корреляционное отношение и индекс корреляции - это числовые характеристики, тесно связанные понятием случайной величины, а точнее с системой случайных величин. Поэтому для введения и определения их значения и роли необходимо пояснить понятие системы случайных величин и некоторые свойства присущие им.

Два или более случайные величины, описывающих некоторое явление называют системой или комплексом случайных величин.

Систему нескольких случайных величин X, Y, Z, …, W принято обозначать через (X, Y, Z, …, W).

Например, точка на плоскости описывается не одной координатой, а двумя, а в пространстве - даже тремя.

Свойства системы нескольких случайных величин не исчерпываются свойствами отдельных случайных величин, входящих в систему, а включают также взаимные связи (зависимости) между случайными величинами. Поэтому при изучении системы случайных величин следует обращать внимание на характер и степень зависимости. Эта зависимость может быть более или менее ярко выраженной, более или менее тесной. А в других случаях случайные величины оказаться практически независимыми.

Случайная величина Y называется независимой от случайной величины X, если закон распределения случайной величины Y не зависит от того какое значение приняла величина X.

Следует отметить, что зависимость и независимость случайных величин есть всегда явление взаимное: если Y не зависит от X, то и величина X не зависит от Y. Учитывая это, можно привести следующее определение независимости случайных величин.

Случайные величины X и Y называются независимыми, если закон распределения каждой из них не зависит от того, какое значение приняла другая. В противном случае величины X и Y называются зависимыми.

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Понятие "зависимости" случайных величин, которым пользуются в теории вероятностей, несколько отличается от обычного понятия "зависимости" величин, которым пользуются в математике. Так, математик под "зависимостью" подразумевает только один тип зависимости - полную, жесткую, так называемую функциональную зависимость. Две величины X и Y называются функционально

зависимыми, если, зная значение одного из них, можно точно определить значение другой.
В теории вероятностей встречаются несколько с иным типом зависимости - вероятностной зависимостью. Если величина Y связана с величиной X вероятностной зависимостью, то, зная значение X, нельзя точно указать значение Y, а можно указать её закон распределения, зависящий от того, какое значение приняла величина X.
Вероятностная зависимость может быть более или менее тесной; по мере увеличения тесноты вероятностной зависимости она все более приближается к функциональной. Т.о., функциональную зависимость можно рассматривать как крайний, предельный случай наиболее тесной вероятностной зависимости. Другой крайний случай - полная независимость случайных величин. Между этими двумя крайними случаями лежат все градации вероятностной зависимости - от самой сильной до самой слабой.
Вероятностная зависимость между случайными величинами часто встречается на практике. Если случайные величины X и Y находятся в вероятностной зависимости, то это не означает, что с изменением величины X величина Y изменяется вполне определенным образом; это лишь означает, что с изменением величины X величина Y имеет тенденцию также изменяться (возрастать или убывать при возрастании X). Эта тенденция соблюдается лишь в общих чертах, а в каждом отдельном случае возможны отступления от неё.
1.2 Числовые характеристики случайных величин
корреляционный случайный величина зависимость
До сих пор мы обсуждали свойства систем случайных величин, давая только словесное разъяснение. Однако существуют числовые характеристики, посредством которых исследуются свойства как отдельных случайных величин, так и системы случайных величин.
Одной из важнейших характеристик случайной величины нормального распределения является математическое ожидание.
Рассмотрим дискретную случайную величину X, имеющую возможные значения X1, X2, ... , Xn с вероятностями p1, p2, ... , pn. нам требуется охарактеризовать каким-то числом положение значений случайной величины на оси абсцисс с учетом того, что эти значения имеют различные значения. Для этой цели обычно пользуются так называемым "средним взвешенным" из значений Xi, причем каждое значение Xi при осреднении должно учитываться с "весом", пропорциональным вероятности этого значения. Таким образом, если обозначить "среднее взвешенное" через M[X] или mx, получим
(1)

или, учитывая, что
, то (2)
Математическим ожиданием случайной величины называется сумма произведений всех возможных значений случайной величины на вероятности этих значений.
Для большей наглядности рассмотрим одну механическую интерпретацию

введенного понятия. Пусть на оси абсцисс расположены точки с абсциссами x1, x2, ..., xn, в которых сосредоточены соответственно массы p1, p2, ... , pn, причем . Тогда математическое ожидание есть не что иное, как абсцисса центра тяжести данной системы материальных точек.
Формула (2) для математического ожидания соответствует случаю дискретной случайной величины. Для непрерывной величины X математическое ожидание, естественно, выражается не суммой, а интегралом:
(3)

где - плотность распределения величины X.
Формула (3) получается из формулы (2), если в ней заменить отдельные значения Xi непрерывно изменяющимся параметром X, соответствующие вероятности pi элементом вероятности f(x)dx, конечную сумму - интегралом. В механической интерпретации математическое ожидание непрерывной случайной величины сохраняет тот же смысл - абсциссы центра тяжести в случае, когда масса распределения по оси абсцисс непрерывна с плотностью f(x). Следует отметить, что математическое ожидание существует не для всех случайных величин, что, однако, по мнению некоторых ученых, не представляет для практики существенного интереса.
Помимо математического ожидания важное значение имеют также другие числовые случайной величины - моменты.
Понятие момента широко применяется в механике для описания распределения масс (статистические моменты, моменты инерции и т.д.). Совершенно теми же приемами пользуются в теории вероятностей для описания основных свойств распределения случайной величины. Чаще всего применяются на практике моменты двух видов: начальные и центральные.
Начальным моментом s-го порядка прерывной случайной величины X называется сумма вида
(4)

Очевидно это определение совпадает с определением начального момента порядка s в механике, если на оси абсцисс в точках x1, ..., xn сосредоточена масса p1, ..., pn.
Для непрерывной случайной величины X начальным моментом s-го порядка называется интеграл

(5)

Очевидно, что
(6)

т.е. начальный момент s-го порядка случайной величины X есть не что иное, как математическое ожидание s-ой степени этой случайной величины.
Перед тем как дать определение центрального момента введем понятие

"центрированной случайной величины".
Пусть имеется случайная величина X с математическим ожиданием mx.
Центрированной случайной величиной, соответствующей величине X, называется отклонение случайной величины X от её математического ожидания
(7)

Нетрудно видеть, что математическое ожидание центрированной случайной величины равно нулю.
Центрирование случайной величины равносильно переносу начала координат в точку, абсцисса которой равна математическому ожиданию.
Центральным моментом порядка s случайной величины X называется математическое ожидание s-ой степени соответствующей центрированной случайной величины:
(8)
Для прерывной случайной величины s-й центральный момент выражается суммой
(9)
а для непрерывной - интегралом
. (10)
Важнейшее значение имеет второй центральный момент, который называют дисперсией и обозначают D[X]. Для дисперсии имеем
. (11)
Дисперсия случайной величины есть характеристика рассеивания, разбросанности значений случайной величины около её математического ожидания. Само слово "дисперсия" означает "рассеивание".
Механической интерпретацией дисперсии является не что иное, как момент инерции заданного распределения масс относительно центра тяжести. На практике часто применяется также величина
, (12)
называемая средним квадратичным отклонением (иначе - "стандартом") случайной величины X.
Теперь перейдем к рассмотрению характеристик систем случайных величин.
Начальным моментом порядка k,s системы (X, Y) называется математическое ожидание произведения Xk и Ys,
xk,s=M[XkYs] (13)
Центральным моментом порядка k,s системы (X, Y) называется математическое ожидание произведения k-ой и s-ой степени соответствующих центрированных величин:
(14)
где , .
Для непрерывных случайных величин
(15)
, (16)
где f(x,y) - плотность распределения системы.

Помимо чисел k и s, характеризующих порядок момента по отношению к отдельным величинам, рассматривается ещё суммарный порядок момента k+s, равный сумме показателей степеней при X и Y. Соответственно суммарному порядку моменты классифицируют на первый, второй и т.д. На практике обычно применяются только первые и вторые моменты.
Первые начальные моменты представляют собой математические ожидания величин X и Y, входящих в систему y1,0=mx, y0,1=my.
Совокупность математических ожиданий mx, my представляет собой характеристику положения системы. Геометрически это координаты средней точки на плоскости, вокруг которой происходит рассеивание точки (X, Y).
Важную роль на практике играют также вторые центральные моменты систем. Два из них представляют собой дисперсии величин X и Y , характеризующие рассеивание случайной точки в направлении осей Ox и Oy. Особую роль играет второй смещенный центральный момент:
(17)
называемый корреляционным моментом (иначе - "моментом связи")случайных величин X и Y.
Корреляционный момент есть характеристика системы случайных величин, описывающая, помимо рассеивания величин X и Y, еще и связь между ними. Для того, чтобы убедиться в этом отметим, что корреляционный момент независимых случайных величин равен нулю.
Заметим, что корреляционный момент характеризует не только зависимость величин, но и их рассеивание. Поэтому для характеристики связи между величинами (X;Y) в чистом виде переходят от момента Kxy к характеристике
(18)
где yx, yy - средние квадратичные отклонения величин X и Y. Эта характеристика называется коэффициентом корреляции величин X и Y.
1.3 Корреляционное отношение

Коэффициент корреляции является полноценным показателем тесноты связи лишь в случае линейной зависимости между переменными. Однако часто возникает необходимость в достоверном показателе интенсивности связи при любой форме зависимости.
Для получения такого показателя вспомним правило сложения дисперсий (19)
где S2y -- общая дисперсия переменной
(20)
S'2iy -- средняя групповых дисперсий Sy , или остаточная дисперсия --
(21)
(22)
(23)
Остаточной дисперсией измеряют ту часть колеблемости Y, которая возникает из-за изменчивости неучтенных факторов, не зависящих от X.
Межгрупповая дисперсия выражает ту часть вариации Y, которая обусловлена

изменчивостью X. Величина
(24)
получила название эмпирического корреляционного отношения Y по X. Чем теснее связь, тем большее влияние на вариацию переменной доказывает изменчивость X по сравнению с неучтенными факторами, тем выше зух.
Величина з2ух, называемая эмпирическим коэффициентом детерминации, показывает, какая часть общей вариации Y обусловлена вариацией X. Аналогично вводится эмпирическое корреляционное отношение X по Y.

(25)
Отметим основные свойства корреляционных отношений (при достаточно большом объеме выборки п):
1. Корреляционное отношение есть неотрицательная величина, не превосходящая 1: 0?з? 1.
2. Если з = 0, то корреляционная связь отсутствует.
3. Если з= 1, то между переменными существует функциональная зависимость.
4. зху? зху т.е. в отличие от коэффициента корреляции r (для которого ryx = rxy = r) при вычислении корреляционного отношения существенно, какую переменную считать независимой, а какую -- зависимой.
Эмпирическое корреляционное отношение зху является показателем рассеяния точек корреляционного поля относительно эмпирической линии регрессии, выражаемой ломаной, соединяющей значения yi . Однако в связи с тем, что закономерное изменение у, нарушается случайными зигзагами ломаной, возникающими вследствие остаточного действия неучтенных факторов, Rxy преувеличивает тесноту связи. Поэтому наряду с зху рассматривается показатель тесноты связи Ryx, характеризующий рассеяние точек корреляционного поля относительно линии регрессии ух .
Показатель Ryx получил название теоретического корреляционного отношения или индекса корреляции Y по X

(26)
где дисперсии д2у и s'y2 определяются по (20) - (22), в которых групповые средние yi, заменены условными средними yxi, вычисленными по уравнению регрессии.
Подобно Ryx вводится и индекс корреляции X по Y
(27)
Достоинством рассмотренных показателей з и R является то, что они могут быть вычислены при любой форме связи между переменными. Хотя з и завышает тесноту связи по сравнению с R, но для его вычисления не нужно знать уравнение регрессии.
Корреляционные отношения з и R связаны с коэффициентом корреляции r следующим образом:
(28)
Покажем, что в случае линейной модели, т.е. зависимости
ух - у = byx (х - х), индекс корреляции Rxy равен коэффициенту корреляции r (по

абсолютной величине): $R_{yx} = |r|$ (или $R_{yx}= |r|$), для простоты $n_i = 1$. По формуле (26)

(так как из уравнения регрессии $y_{xi}-y=b_{yx}(x_i-x)$
Теперь, учитывая формулы дисперсии, коэффициентов регрессии и корреляции, получим:
(29)

1.4 Индекс корреляции

Коэффициент индекса корреляции показывает долю общей вариации зависимой переменной, обусловленной регрессией или изменчивостью объясняющей переменной. Чем ближе индекс корреляции к 1, тем теснее наблюдения примыкают к линии регрессии, тем лучше регрессия описывает зависимость переменных.
Проверка значимости корреляционного отношения з основана на том, что статистика
(30)
(где т -- число интервалов по группировочному признаку) имеет F-распределение Фишера - Снедекора с к1=т-1 и k2=n - т степенями свободы. Поэтому з значимо отличается от нуля, если F >Fa,k1,k2, где F a,k1,k2 - табличное значение F-критерия на уровне значимости б при числе степеней свободы к1= т - 1 и к2= п - т.
Индекс корреляции R двух переменных значим, если значение статистики:
(31)
больше табличного F a,k1,k2, где к1=1 и k2 = n - 2.

1.5 Коррелированность и зависимость случайных величин

Две случайные величины х и у называют коррелированными, если их корреляционный момент (или, что то же, коэффициент корреляции) отличен от нуля; Х и у называют некоррелированными величинами, если их корреляционный момент равен нулю. Две коррелированные величины также и зависимы. Действительно, допустив противное, мы должны заключить, что Кху=0, а это противоречит условию, так как для коррелированных величин Кху?0. Обратное предположение не всегда имеет место, т. е. если две величины зависимы, то они могут быть как коррелированными, так и некоррелированными. Другими словами, корреляционный момент двух зависимых величин может быть не равен нулю, но может и равняться нулю.
Итак, из коррелированности двух случайных величин следует их зависимость, но из зависимости еще не вытекает коррелированность из независимости двух величин следует их некоррелированность, но из некоррелированности еще нельзя заключить о независимости этих величин.

Глава 2. Пример вычисления корреляционного отношения

2.1 Вычисление корреляционного отношения

Закон распределения дискретной двумерной случайной величины задан таблицей:

X / Y

-1

0

1

0

0.15

0.40

0.05

1

0.20

0.10

0.10

Необходимо найти коэффициент корреляции rxy
Находим распределение составляющих X и Y:

X

0

1

p

0.6

0.40

Y

-1

0

1

p

0.35

0.50

0.15

Находим математическое ожидание составляющих:
Mx= 0*0.6+1*0.40=0.4
My= -1*0.35+0*0.50+1*0.15=-0.20
Их можно было бы найти, используя формулу:
Находим дисперсии составляющих:
Стало быть:

Находим Mxy, по формуле
Mxy = 0*(-1)*0.15+0*0*0.40+0*1*0.05+1*(-1)*0.20+1*0*0.10+1*1*0.10=-0.10
Можно было бы составить закон распределения Z=X*Y, а затем найти Mz=Mxy:

Z=X*Y

-1

0

1

p

0.20

0.70

0.10

Mz = -1*0.20+0*0.70+1*0.10=-0.10

Находим корреляционный момент, используя формулу (17) или
Kxy =-0.10-0.40*(-0.20) = - 0.02 ? 0
Находим коэффициент корреляции по формуле (18):
 - отрицательная корреляция.
Заключение

Выполнена курсовая работа по предмету «Теория вероятностей и математическая статистика» на тему «Корреляционное отношение. Индекс корреляции». В данной курсовой работе были рассмотрены следующие вопросы: числовые характеристики

случайных величин, корреляционное отношение, индекс корреляции, коррелированность и зависимость случайных величин. А также был рассмотрен пример вычисления корреляционного отношения и составлена программа, с помощью которой можно найти корреляционное отношение.
В курсовой представлена вся необходимая информация, использована как научная литература, так и возможности всемирной сети Internet.
Список используемой литературы

1) Эконометрика: Начальный курс/Я.Р. Магнус, П.К. Катышев.;
2) Теория вероятностей и математическая статистика: Учебник для вузов.-М.2002.-543с.;
3) Кремер Н.Ш. Математическое программирование.М.1995;
4) Теория вероятностей. Под редакцией д-ра техн. наук, профессора В.С. Зарубина и д-ра физ.-мат. наук, профессора А.П. Крищенко. М.2004;
5) В.Е. Гмурман. Теория вероятностей и математическая статистика. М.2003;
6) Письменный Д.Т. Конспект лекций по теории вероятностей и математической статистике М.2004;

Приложение

Программа для нахождения корреляционного отношения:
1) блок схема2) текст программы
```
program korrel;
uses crt;
var
Mx,Mx1,Qx,My,My1,Qy,Dx,Dy,Mxy,Kxy,Rxy,h,h1,s,s1,f,t:real;
k,e,c,d,i,j:integer;
a:array[1..100,1..100] of real;
X:array [1..100] of real;
X1:array [1..100] of real;
Y:array [1..100] of real;
Y1:array [1..100] of real;
begin
clrScr;
a[1,1]:=0;
write('Vvedite kolichestvo znachenii po X:');
readln(c);
write('Vvedite kolichestvo znachenii po Y:');
readln(d);
writeln('X:');
for k:=2 to c+1 do
readln(a[k,1]);
writeln('Y:');
for e:=2 to d+1 do
```

```
readln(a[1,e]);
writeln('Vvedite otnositelnie chastoti:');
for k:=2 to c+1 do
for e:=2 to d+1 do
readln(a[k,e]);
for k:=1 to c+1 do
begin
for e:=1 to d+1 do
write(' ',a[k,e]:4:2,' ');
writeln;
end;
for k:=2 to c+1 do
begin x1[k-1]:=0;
for e:=2 to d+1 do
begin
x1[k-1]:=x1[k-1]+a[k,e];
end;
end;
for e:=2 to d+1 do
begin y1[e-1]:=0;
for k:=2 to c+1 do
begin y1[e-1]:=y1[e-1]+a[k,e]; end; end;
{for i:=1 to c do
write('',x1[i]:4:2,' ');
writeln;
for i:=1 to d do
write('',y1[i]:4:2,' ');
writeln;}
for i:=1 to c do
begin x[i]:=a[i+1,1]; end;{ write('',x[i]:4:2,' '); end; writeln;}
for j:=1 to d do
begin y[j]:=a[1,j+1]; end; { write('',y[j]:4:2,' '); end; writeln;}
begin
t:=0;
for k:=2 to c+1 do
begin
for e:=2 to d+1 do
begin
Mxy:=a[k,1]*a[1,e]*a[k,e];
t:=t+Mxy;
end;
end;
end;
```

```
begin
h:=0;
for i:=1 to c do
begin
Mx:=x[i]*x1[i];
h:=h+Mx;
end;
h1:=0;
for i:=1 to c do
begin
Mx1:=x[i]*x[i]*x1[i];
h1:=h1+Mx1;
Dx:=h1-sqr(h1);
end;
Qx:=sqrt(Dx);
end;
begin
s:=0;
for i:=1 to d do
begin
My:=y[i]*y1[i];
s:=s+My;
end;
s1:=0;
for i:=1 to d do
begin
My1:=y[i]*y[i]*y1[i];
s1:=s1+My1;
Dy:=s1-sqr(s);
end;
Qy:=sqrt(Dy);
Kxy:=t-h*s;
Rxy:=Kxy/(Qx*Qy);
Writeln('Koeficient korrelyacii=',Rxy:2:2);
end;
readkey;
end.
```

3) пример выполнения программы

```
Vvedite kolichestvo znachenii po X:2
Vvedite kolichestvo znachenii po Y:3
X:
0
1
```

Y:
-1
0
1
Vvedite otnositelnie chastoti:
0.15
0.40
0.05
0.20
0.10
0.10
0.00 -1.00 0.00 1.00
0.00 0.15 0.40 0.05
1.00 0.20 0.10 0.10
Koeficient korrelyacii=-0.06