# Теория вероятностей

## Учебно-методическое пособие для сдачи зачета по курсу 3 семестра.

Данное пособие представляет собой план ответа для зачета у Матвеева В.Ф. (201,202 группы). Приведены в сжатой форме на содержательном уровне все основные определения и теоремы, рассматривавшиеся на семинарских занятиях по курсу «Теория вероятностей»

Draft version.

Часть 1. Вероятностное пространство

Основные понятия.

*Теория вероятностей* позволяет строить математические модели для некоторых экспериментов. Как правило, стохастические методы применяют при создании моделей *случайных экспериментов*, то есть таких, результат которых не детерминирован, которые гипотетически могут быть повторены любое количество раз и при этом обладаю статистической устойчивостью. В соответствии с аксиоматикой Колмогорова, базовой частью модели является вероятностное пространство, в терминах которого решается задача. *Вероятностное пространство* – это тройка ( $\Omega, A, P$ ), где

$\Omega$ - *пространство элементарных исходов* - вообще говоря, любое множество (конечное или бесконечное). Предполагается, что это множество описывает все возможные исходы эксперимента

А – *сигма-алгебра событий*. Событие есть подмножество омега, сигма-алгебра – класс, такой что

$1.\Omega\in A\ 2.\ a\in A\Rightarrow\bar{a}\in A\ 3.\ \text{для любой счетной последовательности}\ a_n: \bigcup_n a_n\in A$ .

Введение этого объекта вполне естественно, поскольку при исследовании нас часто могут интересовать объединения и пересечения событий.

Р – *вероятность*. Это функция-мера, заданная на множестве событий. По определению, $0\leqslant P(A)\leqslant 1$ и функция Р(А) сигма-аддитивна. (Более точно, в аксиомах Колмогорова есть три пункта:

$1.P(A)>0\ \forall A\ 2.P(\Omega)=1\ 3.P(A\cup B)=P(A)+P(B),\ \text{если}\ A\cap B=\emptyset$ ). Для формальной теории не важно, каким образом определена вероятность, в реальности же этот вопрос весьма сложный. Обычно если нет оснований предполагать, что вероятность одного события в силу каких-то факторов больше вероятности другого, то эти вероятности полагают равными.

*Условной вероятностью* А при условии В называется

$P(A|B)=\frac{P(A\cap B)}{P(B)}, P(B)\neq 0$ . События называются *независимыми*, если $P(A\cap B)=P(A)\cdot P(B)$ .

*Формула полной вероятности* $P(A)=\sum_k (P(H_k)\cdot P(A|H_k))$ позволяет найти вероятность события, разбив его на части некоторой группой событий (гипотез). Требуется, чтобы все гипотезы были несовместными (не пересекались) и в сумме покрывали полностью событие А.

*Формула Байеса* $P(H_k|A)=\frac{P(H_k)P(A|H_k)}{\sum_i P(H_i)P(A|H_i)}$ позволяет узнать вероятность гипотезы a posteriori.

Модели вероятностных пространств.

Рассмотрим урну с m белыми и n черными шарами. Из нее наугад вытаскивают один шарик. Пронумеруем шарики от 1 до n+m. Мы считаем, что эксперимент идеальный, и вероятность вытащить каждый из n+m шариков одинаковая.  Элементарными исходами этого эксперимента будут факты появления шарика с тем или иным номером (цвет пока не важен). Пусть множество событий есть множество всех подмножеств множества элементарных исходов. Тогда, так как вероятность события, равного всему множеству элементарных исходов равна 1, а вероятности вытащить каждый отдельный шарик можно считать равными друг другу (эксперимент идеальный), получается, что вероятность вытащить каждый конкретный шарик равна 1/(n+m). Тогда, если вспомнить, что n из шариков черные, в силу аддитивности вероятности получим, что вероятность вытащить черный шар равна n/(n+m).
Эти идеи можно развить, например, рассмотрев несколько последовательных извлечений с возвращением. Тогда элементарными исходами будут упорядоченные последовательности номеров шаров. Из соображений комбинаторики несложно рассчитать вероятности различных событий в такой модели. Конкретные числа указаны в части 3, при описании биномиального распределения.

## Часть 2. Случайные величины

### Определения

*Случайная величина* есть функция от элементарных исходов. Целью является установить соответствие между вероятностным пространством и вещественной прямой, на которой привычнее задавать функции. Поэтому требуется, чтобы отображение $\xi:\Omega\to\mathbb{R}$ было измеримым, то есть на R должна быть задана сигма-алгебра, и полный прообраз любого элемента этой сигма-алгебры должен входить в множество событий. Обычно берется борелевская сигма-алгебра, то есть сигма-алгебра, порожденная всеми открытыми интервалами. Остается установить однозначное соответствие для вероятности. *Распределение вероятностей* случайной величины – это функция B->R, такая, что $\forall\Delta\in B: P(\Delta)=P(\omega:\xi(\omega)\in\Delta)$. Чаще всего удобнее пользоваться *функцией распределения*: $F_\xi(x)=P(\xi<x)\forall x\in\mathbb{R}$
Случайная величина называется *дискретной*, если существует не более чем счетное множество X, такое что P(X)=1. Дискретная случайная величина принимает не более, чем счетное число значений.
Случайная величина называется *непрерывной* (абсолютно непрерыной), если существует неотрицательная функция $p_\xi(t), такая что F_\xi(x)=\int\limits_{-\infty}^{x} p_\xi(t)dt$.
Функция p называется *плотностью распределения*.
Возможно построение функций распределения, не относящихся к описанным двум классам, но такие функции маргинальны.
Во многих исследованиях удобны функции распределений в преобразованном виде: *характеристическая функция* (преобразование Фурье) $f_\xi(t)=Me^{it\xi}$, *производящая функция* $g_\xi(z)=Mz^\xi$ (для дискретных целочисленных случайных величин), *преобразование Лапласа-Стилтьеса* $\varphi(s)=Me^{-s\xi}$ (для неотрицательных случайных величин, по сути аналогично характеристической функции, но не содержит комплексных чисел в явном виде). Эти преобразования однозначно задают распределения. Основное свойство характеристической функции: $f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$, то есть с ее помощью удобно исследовать распределения сумм случайных величин. С

помощью этих преобразований можно находить моменты как производные в некоторых точках.

Характеристики

На практике важны некоторые характеристики случайных величин.
*Математическое ожидание* – показатель центра распределения - по определению $M\xi = \int\limits_{\Omega} \xi(\omega) P(d\omega)$ . В частности, для дискретной случайной величины: $\sum\limits_{k=1}^{+\infty} x_k \cdot p_k$ ,для непрерывной случайной величины $M\xi = \int\limits_{-\infty}^{+\infty} x \cdot p_\xi(x) dx$ . Если ряд или интеграл не сходится абсолютно, то говорят, что математического ожидания не существует. Также показателями центра распределения являются *мода* – наиболее часто встречающееся значение случайной величины и *медиана* – такая точка M, что $P(\xi \le M) \ge \frac{1}{2} \cup P(\xi \ge M) \ge \frac{1}{2}$ .
*Дисперсия* – показатель разброса значений вокруг центра – по определению $D\xi = M|\xi - M\xi|^2$ . *Среднеквадратичное отклонение* $\sigma = \sqrt{(D\xi)}$
Основные свойства математического ожидания и дисперсии следующие:

$M(a\xi + b\eta) = aM\xi + bM\eta$

$M\xi\eta = M\xi \cdot M\eta, \text{ если } \xi \text{ и } \eta \text{ независимы}$

$D(a\xi + b\eta) = a^2 D\xi + b^2 D\eta, \text{ если } \xi \text{ и } \eta \text{ независимы}$

*k-м моментом* случайной величины называется $M\xi^k$ , *k-м центральным моментом* - $M(\xi - M\xi)^k$ . Эти величины могут характеризовать некоторые другие свойства распределения, например величина $\frac{M(\xi - M\xi)^k}{\sigma^3}$ - показатель асимметрии Пирсона, $\frac{M(\xi - M\xi)^4}{\sigma^4}$ - коэффициент эксцесса (резкого увеличения в некоторой точке)

Предельные теоремы

Виды сходимости случайных величин.
Сходимость почти наверное: $\xi_n \xrightarrow{\text{почти наверное}} \xi \Leftrightarrow P(\omega : \xi_n(\omega) \to \xi) = 1$
Сходимость по вероятности: $\xi_n \xrightarrow{P} \xi \Leftrightarrow \forall \varepsilon : \lim\limits_{n \to \infty} P(|\xi_n - \xi| > \varepsilon) = 0$
Сходимость в среднем порядка l: $\xi_n \xrightarrow{l} \xi \Leftrightarrow \lim\limits_{n \to \infty} M|\xi_n - \xi|^l = 0$
Сходимость по распределению:
$\xi_n \underset{d}{\to} \xi \Leftrightarrow F_{\xi_n}(x) \to F_\xi(x) \text{ поточечно, и } F_\xi(x) \text{ непрерывна } \forall x$
Слабая сходимость:
$\xi_n \xrightarrow{w} \xi \Leftrightarrow \lim\limits_{n \to \infty} M\varphi(\xi_n) = M\varphi(\xi) \forall \varphi(x) - \text{непрерывной и ограниченной}$
Следующая схема показывает связи между видами сходимости:


*Закон больших чисел*: среднее арифметическое большого количества независимых одинаково распределенных случайных величин перестает быть случайной и стремится к постоянной.

$$\frac{\sum_{k=1}^{n} \xi_k}{n} \xrightarrow[n \to +\infty]{P} M\xi$$

Существует несколько форм данной теоремы: требуя существование и/или ограниченность дисперсии (а также накладывая на нее некие ограничения), можно доказать, что имеет место сходимость почти наверное (усиленный закон больших чисел)

*Центральная предельная теорема*. Сумма независимых одинаково распределенных случайных величин ведет себя приблизительно, как нормально распределенная случайная величина. Формулировка:

$$P\left(\frac{\sum_{k=1}^{n} \xi_k - M\sum_{k=1}^{n} \xi_k}{\sqrt{(D\sum_{k=1}^{n} \xi_k)}} < x\right) \xrightarrow[n \to +\infty]{d} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{t^2}{2}} dt \simeq N(0,1)$$

*Теорема Пуассона* обосновывает связь между биномиальным и пуассоновским распределениями. Если число испытаний Бернулли очень велико, а вероятность успеха очень мала, причем так, что np -> 1, то такая случайная величина имеет распределение, близкое к распределению Пуассона. $\sum_{k=1}^{n} \xi_k \simeq B(p) \xrightarrow[\substack{n \to +\infty \\ np \to \lambda}]{d} Pois(\lambda)$

*Неравенство Чебышева* $P(|\xi - M\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$ дает оценку для отклонения случайной величины от ее центра.

Часть 3. Примеры случайных величин

В этой части рассмотрены наиболее часто применяющиеся на практике случайные величины. Для каждой случайной величины приведены содержательный смысл, распределение (таблица для дискретных, функция для непрерывных), математическое ожидание, дисперсия, одно из преобразований, и указана связь с другими случайными величинами. Термин «распределение» используется вместо слов «случайная величина, распределенная ...» как общеупотребительный.

Дискретные распределения

*Распределение Бернулли* B(p). Описывает одиночный эксперимент, в котором нас интересует только одно событие (успех) и его противоположное. $Распределение: \left\| \begin{matrix} 0 1 \\ q p \end{matrix} \right\|; M\xi = p; D\xi = q; g_\xi(z) = q + pz$

*Биномиальное распределение* Bi(n,p) есть сумма n бернуллевских случайных величин. Описывает серию последовательных одинаковых испытаний Бернулли. $Распределение: \left\| \begin{matrix} k \\ C_n^k p^k q^{n-k} \end{matrix} \right\|; M\xi = np; D\xi = npq; g_\xi(z) = (1 + p(z-1))^n$

*Геометрическое распределение* G(p) описывает число успехов до первой неудачи в последовательных испытаниях Бернулли (можно рассматривать число испытаний до первой неудачи, эти две случайных величины отличаются на 1).

$Распределение: \left\| \begin{matrix} k \\ q \cdot p^k \end{matrix} \right\| k = 0,1,2\ldots; M\xi = \frac{p}{q}; D\xi = \frac{p}{q^2}; g_\xi(z) = \frac{1-p}{1-pz}$ Непрерынвым аналогом геометрического распределения является экспоненциальное

*Равномерное распределение* Является моделью для простых экспериментов, где все исходы равновероятны (монеты, кубики etc).

$$Распределение \left\| \begin{matrix} k \\ \frac{1}{n} \end{matrix} \right\| k=1..\mathrm{n}; M\xi=\frac{n+1}{2}; D\xi=\frac{n^2-1}{12}$$

*Распределение Пуассона* Описывает число событий, пришедших в пуассоновском потоке (подробнее в 4 части) за единицу времени

$$Распределение: \left\| \begin{matrix} k \\ e^{-\lambda}\frac{\lambda^k}{k!} \end{matrix} \right\|; M\xi=D\xi=\lambda; g_\xi(z)=e^{\lambda(z-1)}$$ Пуассоновское распределение связано с биномиальным в силу теоремы Пуассона

Непрерывные распределения

*Равномерное распределение* Выбор наугад точки на отрезке.

$$Распределение: p_\xi(x)\equiv\frac{1}{b-a}(a<x<b); M\xi=a+\frac{b}{2}; D\xi=\frac{(b-a)^2}{12}$$

*Экспоненциальное распределение*. Описывает время между появлениями двух событий в потоке.

$$Распределение: \exp(\lambda)=1-e^{-\lambda x}; M\xi=\frac{1}{\lambda}; D\xi=\frac{1}{\lambda^2}; \varphi(s)=\frac{\lambda}{\lambda+s}$$

Экспоненциальное распределение является в некотором смысле непрерывным аналогом геометрического.

*Нормальное распределение*. Описывает попадание в цель, также является предельным распределением сумм случайных величин (см. центральную предельную теорему).

$$Распределение: N(a,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x} e^{\frac{-(t-a)^2}{2\sigma^2}}dt; M\xi=a; D\xi=\sigma^2; f_\xi(t)=e^{ita-\frac{\sigma^2t^2}{2}}$$

Часть 4. Примеры моделей для случайных величин

Урновая модель

В главе 1 описано вероятностное пространство для урновой модели. Эту же модель несложно описать на языке случайных величин. Если вытаскивать один шарик, и записывать 0 для белых и 1 для черных, то однократное испытание моделируется случайной величиной Бернулли с p = n/(n+m), при многократном извлечении с возвращением число успехов (черных шаров) является биномиальной случайной величиной, число вытащенных булых шаров до первого черного – геометрически распределенная случайная величина.

Поток редких событий

Рассмотрим некий механизм (или организацию), к которому извне приходят события. Поток событий будем считать ординарным, стационарным и не имеющим последействия. Поясним эти термины. Под *ординарностью* понимается то, что вероятность появления двух событий одновременно очень мала, стремится к нулю. *Стационарность* означает, что свойства потока одинаковы для временных интервалов одинаковой длины. *Отсутствие последействия* (памяти) означает, что на появление событий в какой-то момент времени не влияет, когда до этого приходили события: $P(T<\xi<T+\tau)=P(0<\xi<\tau)$ . Тогда количество событий, пришедших за единицу времени является пуассоновской случайной величиной, потому что единичный интервал можно разбить на n маленьких интервалов, а выбрав n большим, в силу ординарности можно считать, что на каждом маленьком интервале есть случайная величина Бернулли (либо пришло событие, либо нет). При этом количество пришедших событий никак не зависит от времени и от пришедших ранее событий. Параметр $\lambda$ распределения Пуассона называется интенсивностью потока.

Время между приходом двух событий моделируется экспоненциальной случайной величиной в силу тех же соображений о разбиении интервала на маленькие подинтервалы и того факта, что экспоненциальная случайная величина является непрерывным аналогом геометрической