

Медицинская и биологическая кибернетика

УДК 519.254:616-036.22

МЕТОД ОПРЕДЕЛЕНИЯ ЭФФЕКТОВ СОВМЕСТНОГО ДЕЙСТВИЯ ФАКТОРОВ СРЕДЫ В ЭПИДЕМИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ ДЛЯ БИНАРНЫХ ПЕРЕМЕННЫХ

М.Ю. Антомонов

*ГУ «Институт гигиены и медицинской экологии им. А.Н. Марзеева
НАМН Украины»*

Рассмотрены существующие подходы расчета эффектов совместного действия факторов, использующиеся в медико-экологических исследованиях. Оценены недостатки этих подходов, реализованных в теории вероятностей и математической статистике. Предложена оригинальная методика расчета совместного действия факторов, которая позволяет оперировать с качественной информацией, выраженной в бинарном виде. При конструировании итогового выражения использованы подходы формальной логики и теории вероятностей. Полученная функция представляет собой обобщение формулы расчета вероятности независимых событий с учетом их возможной взаимозависимости. Представлен пример использования этого метода.

Ключевые слова: совместное действие факторов, вероятности независимых и взаимозависимых событий, качественные данные, бинарные переменные.

Розглянуто сучасні підходи розрахунку ефектів сумісної дії факторів, які використовуються в медико-екологічних дослідженнях. Оцінено недоліки цих підходів, реалізованих в теорії ймовірностей і математичній статистиці. Запропоновано оригінальну методіку розрахунку спільної дії факторів, яка дозволяє оперувати з якісною інформацією, що виражена у бінарному вигляді. При конструюванні остаточного виразу використано підходи формальної логіки і теорії ймовірностей. Одержана функція є узагальненням формули розрахунку ймовірності незалежних подій з урахуванням їх можливої взаємозалежності. Надано приклад використання цього методу.

Ключові слова: сумісна дія факторів, ймовірності незалежних і взаємозалежних подій, якісні дані, бінарні змінні.

ВВЕДЕНИЕ

В последнее время в прикладных направлениях математического анализа данных все больший приоритет приобретает практическая направленность методов. Например, такое мощное течение как *Data Mining* гораздо в большей степени ориентировано на практическое приложение полученных результатов, чем на следование математическим канонам [1]. Технология *Data Mining* не зависит от строгого соответствия массивов данных

статистическим критериям или математическому описанию функциональных связей между переменными. Объединяя количественный и качественный анализ данных, технология нацелена на получение новых знаний и принятие решений в условиях многомерности данных и неопределенности их структуры.

В дополнение к классическому регрессионному анализу, предполагающему количественный характер переменных и «нормальность» их распределений, в прикладных исследованиях активно применяются логистические регрессионные модели, являющиеся синтезом функционального и вероятностного описания. Для этих моделей могут использоваться переменные любой шкалы измерения: количественные, ранговые, бинарные и даже оцифрованные метки [2, 3].

Корреспондентский анализ, являющийся развитием корреляционного анализа, позволяет устанавливать связи между неколичественными переменными [4].

Метод Multifactor Dimensionality Reduction (MDR) позволяет описывать связи и оценивать тип взаимодействия для бинарных переменных и строить диаграммы по аналогии с кластерным анализом (для которого, как известно, переменные должны быть только количественными) [5, 6].

Таким образом, вполне оправданным и соответствующим современным тенденциям в математической статистике является разработка таких подходов для анализа данных, которые не столько отличались бы строгостью математического описания, сколько являлись бы синтезом классических приемов и были бы направлены на их практическое использование.

ПОСТАНОВКА ЗАДАЧИ

В медико-экологических исследованиях традиционной является задача оценки действия совместного действия совокупности неблагоприятных факторов окружающей среды на те или иные показатели здоровья. Особенно часто такая проблема возникает в эпидемиологических исследованиях при натурной регистрации выраженности тех или иных экологических и социальных характеристик.

Результаты таких исследований в значительной мере основываются на соответствующей математической (статистической) обработке. В свою очередь, математическая обработка определяется типом полученных данных. В прикладной математике данные разграничиваются на четыре основных типа: количественные, ранговые, бинарные и качественные. От принадлежности к этим типам зависит выбор адекватного математического аппарата [7].

Если и воздействующие факторы $\{x_i\}$ и показатели здоровья $\{y_j\}$ относятся к количественным переменным, наиболее популярными математическими методами являются корреляционный и (или) регрессионный анализ.

В рамках корреляционного анализа, как правило, рассчитываются коэффициенты парной корреляции $r(x_i y_j)$. Их достоверные значения

свидетельствуют в пользу причинно-следственных связей для некоторой (устраивающей исследователя) совокупности пар: отдельный i -ый фактор среды — отдельный j -ый показатель здоровья. Значительно реже применяется расчет множественных коэффициентов корреляции, что более правильно для выявления эффектов совместного действия факторов. Но, в любом случае, выделение с помощью корреляционного анализа в результирующей реакции биосистемы составляющих, связанных именно с совместным действием факторов, крайне затруднено.

Более адекватным приемом является построение регрессионных функций, описывающих зависимость исследуемых характеристик здоровья от совокупности всех учитываемых (регистрируемых) факторов окружающей среды. Эффекты совместного действия в уравнениях множественной регрессии, как правило, связываются с коэффициентами при некоторых искусственных свертках исходных переменных (чаще всего для свертки используется произведение переменных). При этом появляется возможность расчета прогнозных значений здоровья и относительного вклада каждого из исследуемых факторов.

В том случае, если и воздействующие факторы, и показатели здоровья получены в ранговой шкале, адекватными методами являются ранговый корреляционный анализ или составление таблиц сопряженности.

Если среда оценивается в бинарном виде, например, как «опытный» («загрязненный») и «контрольный» («условно чистый») районы, а показатели здоровья относятся к количественной шкале измерений, то традиционно применяется критерий Стьюдента. С его помощью устанавливаются различия между «чистой» и «грязной» территорией для показателей здоровья, на основании чего делается вывод о влиянии на здоровье одного фактора, обуславливающего «загрязнение», или их совокупности.

Часто эффект изменения (ухудшения) здоровья регистрируется в бинарном виде («изменение есть / изменения нет»), а воздействующие факторы — в количественной шкале. В таком случае целесообразно применение дискриминантного анализа, с помощью которого строятся классификационные функции и рассчитывается вероятность ухудшения здоровья [8, 9].

Однако в практике медико-экологических исследований не всегда есть возможность получения количественной информации о загрязнении окружающей среды. Иногда информация носит качественный характер констатации самого факта загрязнения объекта исследования типа «да» — «нет». Кроме того, в ходе математического анализа данных часто возникает потребность преобразования количественных данных (если они есть) в бинарную качественную шкалу. Бинарные переменные легко преобразуются в вероятности (проценты). Иногда и сами показатели статистической отчетности также фиксируются в процентах фактов загрязнения («процент проб, превышающих гигиенический норматив»).

Таким образом, задача описания результатов очень часто попадает в область применения теории вероятностей.

Цель — разработка методики определения совместного действия факторов, позволяющей оперировать с качественной (бинарной) информацией и использующей подходы и методы теории вероятностей.

СОСТОЯНИЕ ПРОБЛЕМЫ

Рассмотрим ситуацию, когда данные и о среде, и о здоровье получены либо в качественном виде («чистый» и «грязный» район и «хорошее» и «плохое» здоровье), либо преобразованы к такому виду, исходя из целей исследования.

Традиционным и широко используемым приемом обработки такого типа данных является составление четырехпольных таблиц сопряженности с последующим подсчетом соответствующих статистических критериев различия (связности), например, χ^2 Пирсона.

В последнее время популярным методом обработки четырехпольных таблиц в эпидемиологии становится расчет относительного риска (RR), позволяющий оценивать «степень вредности воздействия» в относительных единицах.

Однако и в первом, и во втором варианте корректно рассчитываются характеристики связности только для пар «воздействие – эффект». Для оценки совместного действия совокупности факторов приходится прибегать к вспомогательным расчетам, являющимся не всегда корректными.

МЕТОД ОПРЕДЕЛЕНИЯ СОВМЕСТНОГО ДЕЙСТВИЯ ФАКТОРОВ

Рассмотрим возможность использования имеющихся в классической теории вероятностей методов и формул, с помощью которых можно рассчитать совместные эффекты действия [10].

Для согласования строгих подходов теории вероятностей и предлагаемых прикладных приемов сделаем следующие допущения. Во-первых, мы не всегда будем соблюдать аксиоматику этой математической дисциплины, делая упор на утилитарном использовании основных математических формул. Во-вторых, определимся с понятием «событие» — ключевым понятием для расчета соответствующих вероятностей.

Будем считать, что событием x_{jy} является факт одновременного наличия и действующего i -го фактора, и эффекта от его действия (т.е. заболевания $-y$) при j -ом измерении (регистрации медико-экологической ситуации в j -ом пункте). Если такое событие происходит, $x_{jy} = 1$, в противном случае $x_{jy} = 0$.

Тогда вероятность действия i -го фактора на показатель здоровья y будет равна:

$$P_{iy} = \frac{1}{m} \sum_{j=1}^m x_{jy} ,$$

где m — число точек регистрации.

Рассмотрим несколько вариантов расчета вероятности совместного действия разных факторов.

1. Наиболее простой для нашей цели в теории вероятностей является формула расчета *вероятности совместного действия независимых событий*. В соответствии с ней вероятность P_{Xy} совместного действия независимых событий $\{x_i\} = X$ на произвольную характеристику здоровья (y) будет выглядеть следующим образом:

$$P_{Xy} = 1 - (1 - p_{1y})(1 - p_{2y}) \dots (1 - p_{ny}) = 1 - \prod_{i=1}^n (1 - p_{iy}), \quad (1)$$

где n — количество действующих факторов.

Главные недостатки применения этой формулы в медико-экологических исследованиях следующие.

Само предположение о независимости действующих факторов вряд ли оправдано, поскольку, как правило, объекты исследования, значительно «загрязненные» одним из факторов, оказываются «загрязненными» и другими факторами. Это приводит к их достоверной множественной корреляции, что неоднократно выявлялось на анализируемых массивах данных.

Формула очень чувствительна к количеству факторов. Если имеются незначительные частные вероятности возникновения эффектов, только слегка (и недостоверно) превышающие неопределенность (например, 51%), то уже для $n = 3$ итоговая вероятность приближается к 90%, а при $n = 5$ достигает 97%. При больших частных вероятностях, даже при учете двух-трех факторов, вероятность эффекта становится практически стопроцентной.

2. Другим вариантом расчета совместного действия является привлечение понятия «условной вероятности». В таком случае событием — причиной, можно считать одновременное действие всех n факторов, а возникновение результата (например, заболевания) считать зависимым от этой причины.

Вероятность одновременного действия n факторов подсчитывается как отношение суммы событий наличия всех действующих факторов x_n к числу измерений m :

$$P(X) = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n x_{ji} = \frac{x_n}{m}.$$

При этом суммарное $x_n = 1$, если все $x_i = 1$ (логическое перемножение).

Вероятность совместного проявления и совокупности факторов, и наступления эффекта (результата) запишется аналогичным образом:

$$P(Xy) = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n x_{jly} = \frac{x_{ny}}{m}.$$

В этой формуле также $x_{ny} = 1$, если все $x_{iy} = 1$.

Тогда условная вероятность $P(y|X)$ в соответствии с известной формулой теории вероятностей будет равна:

$$P(y|X) = \frac{P(Xy)}{P(X)}. \quad (2)$$

Недостаток использования такого подхода тот же, что и для предложенного ранее варианта, а именно — предположение независимости действия факторов, т.е. отсутствие в формулах эффектов взаимодействия факторов. Следовательно, можно предположить, что рассчитанная таким образом условная вероятность будет меньше, чем гипотетическая «правильная» вероятность, которая должна быть дополнена добавкой, вызванной эффектами взаимодействия.

3. Если изначально предположить взаимозависимость (совместность) эффектов действия различных факторов, то можно воспользоваться формулами произведения событий.

При действии двух факторов формула будет иметь вид:

$$R_{\Sigma y} = p_{1y} + p_{2y} - p_{1y+2y}, \quad (3)$$

где p_{1y+2y} — означает вероятность либо события p_{1y} , либо события p_{2y} (вероятность суммы двух событий или логическое сложение).

Для трех факторов имеем:

$$R_{\Sigma y} = p_{1y} + p_{2y} + p_{3y} - p_{1y+2y} - p_{1y+3y} - p_{2y+3y} + p_{1y+2y+3y}. \quad (4)$$

Недостатком использования такого подхода является значительное усложнение формулы для большего количества факторов, поскольку в нее входят вероятности сумм двух, трех и т.д. событий.

Таким образом, приходим к выводу, что использовать непосредственно формулы (1), (2) или (3) либо некорректно, либо слишком сложно.

В связи с этим нами была предпринята попытка модернизации этих формул. При этом будем основываться на формуле (1), но дополним ее, используя следующие соображения (постулаты).

1. Увеличение вероятности $R_{\Sigma y}$ при увеличении количества сомножителей (факторов) n должно происходить более инерционно, чем в исходной формуле.

2. В формуле должна учитываться возможная связь факторов между собой. При наличии такой связи $R_{\Sigma y}$, рассчитанная по классической формуле, должна уменьшаться, поскольку совместное действие «выбирает на себя» часть от совокупности эффектов независимого действия. При отсутствии совместных взаимодействий факторов формула должна приобретать традиционный вид.

3. Поскольку мы предположили, что исходные данные имеют бинарный вид, мерой такой связности могут быть условные межфакторные вероятности p_{ik} наличия i -го фактора в присутствии k -го фактора (где $i \neq k$).

4. Общая вероятность не должна быть меньше максимальной частной вероятности. То есть, если «наиболее вредный» k -ый фактор вызывает ухудшение здоровья с вероятностью $p_k = p_{max}$ то $P_{\Sigma y} > p_{max}$.

При конструировании итоговой формулы будем использовать подходы формальной логики и теории вероятностей [11, 12].

А именно, если p_{iy} — вероятность действия i -го фактора на показатель здоровья y , то его дополнение $(1 - p_{iy})$ — вероятность «недействия». Произведение вероятностей событий означает вероятность их совместного наступления.

Например, в «классической» формуле

$$P_{\Sigma y} = 1 - \prod_{i=1, \dots, n} (1 - p_{iy})$$

каждый из сомножителей $(1 - p_{iy})$ означает «недействие» i -го фактора, их произведение — *совместное «недействие»*, а дополнение к произведению — *вероятность совместного действия*.

Вероятность совместного наличия любого из факторов i и максимально действующего k запишется как p_{ik} , где $i=1, \dots, k, \dots, n$. Его дополнение $(1 - p_{ik})$ имеет смысл вероятности «отсутствия совместного наличия этих факторов». Произведение этой скобки на p_{iy} , т.е. $p_{iy}(1 - p_{ik})$, означает одновременное *действие* i -го фактора на показатели здоровья (заболеваемость) и *независимость* любого из i -ых факторов от k -го фактора. Соответственно, дополнение к этому произведению $(1 - p_{iy}(1 - p_{ik}))$ имеет смысл «недействия» i -го фактора с учетом его связности с k -ым фактором, т.е. «*совместного зависимого недействия*» факторов i и k .

Далее, перемножая эти скобки для всех факторов, получаем *вероятность «совместного зависимого недействия» для некоторой совокупности воздействующих факторов*. Выделим максимально действующий фактор, и скобку с его «недействием» вынесем за знак произведения.

Тогда *совместное недействие всех факторов*, включая максимально «вредный», запишется в следующем виде:

$$(1 - p_{ky}) \prod_{i=1, i \neq k}^{n-1} (1 - p_{iy}(1 - p_{ik})).$$

Соответственно, дополнение к этому выражению будет иметь смысл вероятности *совместного действия всех факторов с учетом их связи с наиболее действующим, то есть в предположении их зависимого действия*.

Итак, получаем итоговую формулу:

$$P_{\Sigma y} = 1 - (1 - p_{ky}) \prod_{i=1, i \neq k}^{n-1} (1 - p_{iy}(1 - p_{ik})). \quad (5)$$

Очевидно, что при наличии полной независимости всех факторов от ведущего, т.е. при $p_{ik} = 0$, формула приобретает «классический» вид (1). Если хоть одна из вероятностей $p_{ik} \neq 0$, итоговая вероятность по (5) будет меньше вероятности, рассчитанной по (1). При условии полной зависимости всех остальных факторов от «ведущего» ($p_{ik} = 1$), вероятность их «совместного действия» определяется только одним «ведущим» фактором ($P_{\Sigma y} = p_{ky}$).

ПРИМЕР РАСЧЕТА СОВМЕСТНОГО ДЕЙСТВИЯ НЕСКОЛЬКИХ ФАКТОРОВ

Пусть в девяти пунктах наблюдения регистрировалось наличие трех ($n = 3$) вредных факторов среды (x_1, x_2, x_3) и одного показателя здоровья (y). Выраженность переменных кодировалась в бинарном виде: «1» — значимая вредность факторов (например, превышение нормативного значения) или достоверное ухудшение здоровья (например, по сравнению со среднестатистическими значениями); 0 — отсутствие вредности факторов и ухудшения здоровья (табл. 1).

Таблица 1.

Бинарное представление медико-экологической информации

N пункта наблюдения	Факторы среды			Показатель здоровья (y)
	x_1	x_2	x_3	
1	1	1	1	1
2	0	1	1	0
3	1	1	1	1
4	1	1	1	1
5	1	0	1	1
6	1	1	0	1
7	1	0	1	1
8	1	0	1	0
9	1	1	1	0

Событием для расчета вероятностей будем считать совпадения «1» для каждой переменной в сравниваемых столбцах. Вероятность рассчитывается как отношение сумм этих совпадений к количеству замеров (в данном случае, пунктов наблюдения).

Тогда, в соответствии с таблицей число совпадений «1» для факторов x_1 и x_2 равно 5. Соответственно, $p_{12} = 5/9 = 0,56$.

Вероятности связи наличия факторов и ухудшения здоровья будут равны: $p_{1y} = 0,67$, $p_{2y} = 0,44$, $p_{3y} = 0,56$. Максимальной будет вероятность действия первого фактора $p_{1y} = p_{max} = 0,67$, поэтому назовем его «ведущим» фактором (p_{ky}). Вероятности одновременного наличия в пунктах наблюдения этого фактора и двух других равны соответственно: $p_{12} = 0,56$, $p_{13} = 0,78$.

Вероятность совместного действия всех трех факторов на здоровье по предлагаемой формуле (5) будет равна:

$$\begin{aligned}
 P_{\Sigma} &= 1 - (1 - 0,67)[1 - (1 - 0,56) * 0,44] * [1 - (1 - 0,78) * 0,56] = \\
 &= 1 - 0,33 * (1 - 0,44 * 0,44) * (1 - 0,22 * 0,56) = 1 - 0,33 * 0,81 * 0,88 = \\
 &= 1 - 0,14 = 0,76.
 \end{aligned}$$

Для сравнения: по формуле для независимых событий (1) вероятность равна $P_{\Sigma y} = 0,92$, по формуле для условной вероятности (2) — $P_{\Sigma y} = 0,75$, по формуле (3) — $P_{\Sigma y} = 0,33$.

Видно, что вероятность, рассчитанная по формуле (5), находится в интервале между вероятностью изолированного действия наиболее существенного из факторов (нижняя граница) и совместной вероятностью независимого действия (1) всех факторов (верхняя граница), что отвечает сформулированным изначально условиям. Формулы (2) и (3) дают неправильные и явно заниженные результаты, поскольку вероятности, рассчитанные по ним, меньше, чем вероятность изолированного действия наиболее вредного фактора.

Выводы

Показана неадекватность применения методов оценки вероятности совместного действия факторов, используемых в теории вероятностей и математической статистике.

Предложена оригинальная формула расчета вероятности совместного действия факторов, учитывающая их возможную взаимосвязь. При наличии полной независимости всех факторов от ведущего, формула приобретает «классический» вид для расчета совместной вероятности независимых событий. Если имеется хоть одна взаимосвязь, вероятность, рассчитанная по предлагаемой формуле, будет меньше «классической». При условии полной зависимости всех остальных факторов от ведущего, вероятность их совместного действия определяется только максимальной вероятностью эффекта.

1. Дюк В. А., Самойленко А. П. Data Mining. Учебный курс — СПб, 2001. — 368 с.
2. David W. Hosmer, Stanley Lemeshow Applied Logistic Regression, 2nd ed. — New York, Chichester, Wiley, 2002. — 383p.
3. Наследов А. Н. SPSS 19: Профессиональный статистический анализ данных — СПб, 2011. — 400 с.
4. Greenacre M. Correspondence Analysis in Practice, 2nd ed. — London: Chapman & Hall/CRC, — 2007. — 280 p.
5. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. / M.D. Ritchie, L.W. Hahn, N. Roodi & others // Am J Hum Genet. — 2001 Jul;69(1):138-47.
6. Hahn L.W., Ritchie M.D., Moore J.H. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions//Bioinformatics. — 2003 Feb 12;19(3):376-82
7. Орлов А. И. Прикладная статистика — М.: Изд-во «Экзамен» — 2004. — 656 с.
8. Антомонов М.Ю. Математическая обработка и анализ медико-биологических данных — К.: Изд-во „Малый друк“, — 2006. — 558с.
9. Гайдышев И. Анализ и обработка данных — СПб. — 2001. — 750 с.
10. Вентцель Е. С. Теория вероятностей. 10-е изд., стер.. — М.: «Академия», — 2005. — 576 с.
11. Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей — 1970. — 168с.

THE METHOD OF DETERMINATION OF ENVIRONMENTAL FACTORS JOINT IMPACT IN EPIDEMIOLOGICAL STUDIES FOR BINARY DATA

M.Yu. Antomonov

*State Institution "O.N. Marzeev Institute for Hygiene and Medical Ecology
of NAMS of Ukraine" (Kiev)*

Introduction. Modern approaches for data analysis combine classical methods and focused on their practical application. Sometimes the information is presented in the form of qualitative characteristics that are characterize the contamination of the research object. Such binary variables are easily transformed into a probability (in percent), so the task description of results performed using probability theory.

The purpose of the article is to develop such a common method for calculation joint action of the factors that would allow to operate with qualitative (binary) information and would use techniques and formulas of probability theory

Results. A careful analysis was carried out for the existing approaches in the medical and environmental studies for calculating the effect of the joint action of the factors. It was evaluated disadvantages of these approaches that implemented in the theory of probability and mathematical statistics.

The article proposes an original method of calculating the combined effect of the factors that allows you to work with the information expressed in binary form. The final expression was designed by using approach of formal logic and probability theory.

Conclusions. It is shown that the known methods of probability theory cannot be adequately used to evaluate the combined effect of the factors. The original method of calculating the probability of the joint action of factors that take into account their possible connection is described.

Keywords: qualitative data, binary variables, joint effect of the factors, the probability of independent and interdependent events.

1. Duke V. A. Samoilenko A. P. Data Mining. Training — SPb, 2001. — 368p.
2. David W. Hosmer, Stanley Lemeshow Applied Logistic Regression, 2nd ed. — New York, Chichester, Wiley, 2002. — 383p.
3. Nasledov A. N. SPSS 19: Professional statistical analysis. — SPb, 2011. — 400 p.
4. Greenacre M. *Correspondence Analysis in Practice, 2nd ed.* — London: Chapman & Hall / CRC — 2007. — 280 p.
5. Ritchie M. D., Hahn L. W., Roodi N., Bailey L. R., Dupont W. D., Parl F. F., Moore J. H. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* 2001 Jul; 69 (1): 138-47.
6. Hahn L. W., Ritchie M. D., Moore J. H. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions//*Bioinformatics.* 2003 Feb 12; 19 (3): 376-82
7. Orlov A. I. Applied Statistics — М.: Publisher "Exam", 2004. — 656 p.
8. Antomonov M. Y. Mathematical processing and analysis of medical-biological data. — Kiev: Publishing house "Malii Druk", 2006. — 558p.

9. Gaydyshev I. Analysis and data processing — St. Petersburg, 2001. — 750p.
10. Wentzel E. S. Probability 10th ed., Sr — M.: "Academy", 2005. — 576 p.
11. Gnedenko B. V., Khinchin A. Y. An elementary introduction to the theory of probability, 1970. — 168p.
12. Novikov P. S. Elements of mathematical logic. 2nd ed. — M.: Nauka, 1973 — 400 p.

Получено 02.06.2015