

СЛУЧАЙНО-МНОЖЕСТВЕННЫЕ РАЗЛОЖЕНИЯ ДВУДОЛЬНЫХ СЛУЧАЙНЫХ ВЕКТОРОВ

О.Ю. Воробьев, Е.Е. Голденок, Д.В. Семенова *

Рассматриваются двудольные случайные вектора, компоненты которых конструируются из произвольных случайных величин добавлением атома в нуле. Доказаны теоремы, связанные с разложением распределения двудольного случайного вектора по случайно-множественному базису.

Рассмотрим произвольную случайную величину ρ с функцией распределения $F(u)$, непрерывной в нуле, и конечным математическим ожиданием:

$$E\rho = \int_{-\infty}^{\infty} u dF(u) < \infty.$$

Добавим случайной величине ρ "атом" в нуле с вероятностью $1-p$, сохранив ей прежнее распределение с вероятностью p , и обозначим полученную случайную величину $\varrho(p)$. Эта величина имеет функцию распределения (рис. 1):

$$F_p(u) = \begin{cases} pF(u), & u < 0, \\ 1 - p + pF(u), & \text{иначе.} \end{cases}$$

Таким образом, можно сказать, что случайная величина $\varrho(p)$ принимает два *обобщенных значения*, из которых одно — обычный нуль с вероятностью $1-p$, а второе — случайная величина ρ с вероятностью p :

$$\varrho(p) = \begin{cases} 0, & \text{с вероятностью } 1-p, \\ \rho, & \text{с вероятностью } p. \end{cases}$$

Более того, поскольку функция распределения F непрерывна в нуле, то случайная величина ρ не имеет атома в нуле и можно записать, что

$$P(\varrho(p) \neq 0) = p, \quad P(\varrho(p) = 0) = 1 - p.$$

Определение (двудольная случайная величина, или обобщенная бернуллиевская случайная величина). *Двудольной случайной величиной с параметрами (p, ρ) будем называть случайную величину $\varrho(p)$, с вероятностью p имеющую непрерывную в нуле функцию распределения, а с вероятностью $1-p$ — атом в нуле. Случайную величину ρ с непрерывной в нуле функцией распределения будем называть *ненулевым значением двудольной случайной величины $\varrho(p)$.**

Рассмотрим конечное множество событий \mathfrak{X} и "занумерованные" этими событиями двудольные случайные величины, образующие вектор

$$\varrho(p) = \{ \varrho_x(p_x), x \in \mathfrak{X} \}$$

* © О.Ю. Воробьев, Е.Е. Голденок, Д.В. Семенова, Красноярский государственный университет, 2003

с совместной функцией распределения

$$F_{\mathbf{p}}(u_x, \cdot, x \in \mathfrak{X}) = \mathbf{P} \left\{ \bigcap_{x \in \mathfrak{X}} \{ \varrho_x(p_x) < u_x \} \right\}, \quad (1)$$

где $\mathbf{p} = \{p_x, x \in \mathfrak{X}\}$, и индивидуальными функциями распределения

$$F_{p_x}(u_x) = \mathbf{P} \{ \varrho_x(p_x) < u_x \}, \quad x \in \mathfrak{X},$$

соответственно.

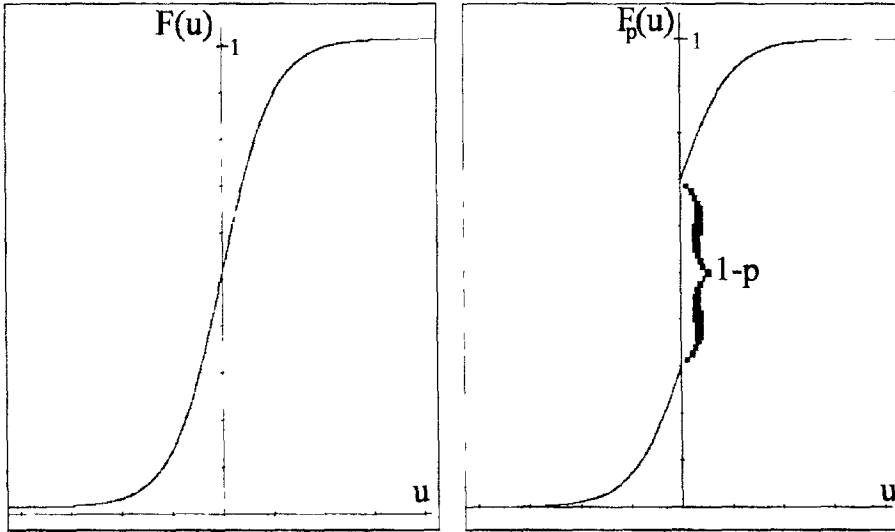


Рис. 1. Функции распределения: слева - случайной величины ρ , справа - двудольной случайной величины $\varrho(p)$

Рассмотрим распределение вероятностей случайного множества K под \mathfrak{X} :

$$p(X) = \mathbf{P}(K = X), \quad X \in 2^{\mathfrak{X}},$$

которое определяется совместной функцией распределения по формулам

$$p(X) = \mathbf{P} \left\{ \bigcap_{x \in X} \{ \varrho_x(p_x) \neq 0 \} \bigcap_{x \in X^c} \{ \varrho_x(p_x) = 0 \} \right\}, \quad X \subseteq \mathfrak{X}, \quad (2)$$

где параметры распределения p_x имеют очевидную интерпретацию

$$p_x = \mathbf{P}(\varrho_x(p_x) \neq 0), \quad x \in \mathfrak{X}$$

как вероятность того, что $\varrho_x(p_x) \neq 0$, а сама вероятность $p(X)$ — это вероятность того, что произошло событие

$$e_X = \left\{ \bigcap_{x \in X} \{ \varrho_x(p_x) \neq 0 \} \bigcap_{x \in X^c} \{ \varrho_x(p_x) = 0 \} \right\},$$

которое заключается в том, что ненулевые компоненты двудольного случайного вектора $\varrho(\mathbf{p})$ образуют подмножество $\{\varrho_x, x \in X\}$. Очевидно, что

$$\mathbf{P}(e_X) = p(X).$$

Определение (двудольный случайный вектор). Вектор $\varrho(\mathbf{p})$, составленный из двудольных случайных величин, определяемых совместной функцией распределения (1), называется *двудольным случайным вектором*.

Определение (случайное множество ненулевых компонент двудольного случайного вектора). Случайное множество K с распределением вероятностей (2) называется случайным множеством ненулевых компонент двудольного случайного вектора.

Утверждение 1. Для любых $X \subseteq \mathfrak{X}$ и $Y \subseteq \mathfrak{X}$, таких что $X \neq Y$ события e_X и e_Y несовместны.

Доказательство. Событие e_X заключается в том, что все компоненты случайного вектора $\varrho(\mathbf{p})$ разбиваются на две части:

$$\{\varrho_x(p_x), x \in \mathfrak{X}\} = \{\varrho_x(p_x) \neq 0, x \in X\} + \{0, x \in X^C\}.$$

Аналогично, событие e_Y

$$\{\varrho_x(p_x), x \in \mathfrak{X}\} = \{\varrho_x(p_x) \neq 0, x \in Y\} + \{0, x \in Y^C\}.$$

То есть имеем два разбиения случайного вектора $\varrho(\mathbf{p})$.

Допустим, что события совместны.

Если $X \neq Y$, то существуют такие $w \in \mathfrak{X}$, что $w \in X$ и $w \in Y^C$. Если $w \in X$, то $\varrho_w(p_w) \neq 0$. С другой стороны, если $w \in Y^C$, то $\varrho_w(p_w) = 0$. Получили противоречие, следовательно, события e_X и e_Y не совместны. Утверждение доказано.

Утверждение 2. Совокупность событий $\{e_X, X \in 2^{\mathfrak{X}}\}$ образует полную группу событий.

Доказательство. По определению совокупность событий e_1, \dots, e_N образует полную группу, если они попарно несовместны и их сумма составляет достоверное событие.

Попарная несовместность событий $\{e_X, X \in 2^{\mathfrak{X}}\}$ следует из утверждения 2.

Стоит заметить, что разбиений всех компонент двудольного случайного вектора на две доли ровно столько, сколько подмножеств имеет конечное множество \mathfrak{X} , т.е. $2^{\mathfrak{X}}$, таким образом, их сумма составляет достоверное событие:

$$\sum_{X \in 2^{\mathfrak{X}}} e_X = \Omega.$$

Следовательно, $\{e_X, X \in 2^{\mathfrak{X}}\}$ образуют полную группу событий. Утверждение доказано.

Теорема (о разложении распределения двудольного случайного вектора по случайно-множественному базису). Пусть $\varrho(\mathbf{p}) = \{\varrho_x(p_x), x \in \mathfrak{X}\}$ — двудольный случайный вектор, составленный из $N = |\mathfrak{X}|$ двудольных случайных величин, с совместной N -мерной функцией распределения

$$F_{\mathbf{p}}(u_x, x \in \mathfrak{X}) = \mathbf{P} \left\{ \bigcap_{x \in \mathfrak{X}} \{\varrho_x(p_x) < u_x\} \right\}.$$

Тогда для функции $F_{\mathbf{p}}(u_x, x \in \mathfrak{X})$ справедливо разложение

$$F_{\mathbf{p}}(u_x, x \in \mathfrak{X}) = \sum_{X \in 2^{\mathfrak{X}}} F_X(u_x, x \in \mathfrak{X}) \cdot p(X), \quad (3)$$

где

$$F_X(u_x, x \in \mathfrak{X} | e_X) = \mathbf{P} \left\{ \bigcap_{x \in \mathfrak{X}} \{ \varrho_x(p_x) < u_x | e_X \} \right\}$$

— условные функции распределения двумерного случайного вектора $\varrho(\mathbf{p})$ при условии случайного события

$$e_X = \left\{ \bigcap_{x \in X} \{ \varrho_x(p_x) \neq 0 \} \bigcap_{x \in X^c} \{ \varrho_x(p_x) = 0 \} \right\}, \quad X \subseteq \mathfrak{X},$$

а

$$p(X) = \mathbf{P}(e_X), \quad X \subseteq \mathfrak{X}$$

— вероятность такого события.

Доказательство. Рассмотрим событие

$$A = \left\{ \bigcap_{x \in \mathfrak{X}} \{ \varrho_x(p_x) < u_x \} \right\}.$$

В силу свойств операций над событиями и утверждения 2 имеем

$$A = A\Omega = A \sum_{X \subseteq 2^{\mathfrak{X}}} e_X = \sum_{X \subseteq 2^{\mathfrak{X}}} Ae_X.$$

Поскольку события e_X попарно несовместны, то и события Ae_X тоже попарно несовместны. Тогда

$$\mathbf{P}(A) = \mathbf{P} \left(\sum_{X \subseteq 2^{\mathfrak{X}}} Ae_X \right) = \sum_{X \subseteq 2^{\mathfrak{X}}} \mathbf{P}(Ae_X).$$

Теперь применим к слагаемым $\mathbf{P}(Ae_X)$ формулу умножения вероятностей:

$$\mathbf{P}(Ae_X) = \mathbf{P}(A | e_X) \mathbf{P}(e_X).$$

Получаем:

$$\mathbf{P} \left\{ \bigcap_{x \in \mathfrak{X}} \{ \varrho_x(p_x) < u_x \} \right\} = \sum_{X \subseteq 2^{\mathfrak{X}}} \mathbf{P} \left\{ \bigcap_{x \in \mathfrak{X}} \{ \varrho_x(p_x) < u_x | e_X \} \right\} \cdot p(X).$$

Теорема доказана.

Определение. Нульмерная функция распределения

$$F_{\emptyset}(\cdot) = \mathbf{P} \left\{ \cdot \mid \emptyset \right\} = \prod_{x \in \mathfrak{X}} \mathbf{1}_{\{u_x \geq 0\}}$$

— это функция распределения вырожденного случайного вектора, который принимает единственное нулевое значение: $\{0, \dots, 0\}$, она равна единице всякий раз, когда все u_x неотрицательны.

Определение. Совокупность условных функций распределения

$$\{F_X(u_x, x \in \mathfrak{X}|e_X), X \subseteq \mathfrak{X}\}$$

образуют *количественную надстройку*.

Таким образом, можно говорить о двухуровневой структуре зависимостей. Первый случайно-множественный уровень отвечает за полную структуру статистических зависимостей и взаимодействий случайных событий, образует случайно-множественный базис, а второй количественный уровень — за структуру зависимостей и взаимодействий компонент двудольного случайного вектора в количественной надстройке.

Основная сложность применения данной теоремы на практике заключается в корректном задании количественной надстройки. Рассмотрим это на примере разложения двумерного двудольного случайного вектора под дуплетом.

Пример разложения распределения двудольного случайного вектора под дуплетом. Рассмотрим $\varrho(\mathbf{p}) = \{\varrho_x, \varrho_y\}$ — двумерный двудольный случайный вектор.

Пусть известно распределение случайно-множественного базиса, то есть распределение случайного множества событий K , определенного под дуплетом событий $\mathfrak{X} = \{x, y\}$ с распределением вероятностей:

$$p(\emptyset) = \mathbf{P}\{\varrho_x = 0, \varrho_y = 0\};$$

$$p(x) = \mathbf{P}\{\varrho_x \neq 0, \varrho_y = 0\};$$

$$p(y) = \mathbf{P}\{\varrho_x = 0, \varrho_y \neq 0\};$$

$$p(xy) = \mathbf{P}\{\varrho_x \neq 0, \varrho_y \neq 0\}$$

распределение случайно-множественного базиса. И пусть известны условные распределения двудольного случайного вектора $\varrho(\mathbf{p})$ при условии соответствующих случайных событий:

$$\begin{aligned} p_\emptyset(u_x, u_y) &= \mathbf{P}\left(\varrho_x = u_x, \varrho_y = u_y \Big|_{e_\emptyset}\right) = \\ &= \begin{cases} 1, & \text{если } u_x = 0, u_y = 0, \\ 0, & \text{иначе} \end{cases} ; \\ p_x(u_x, u_y) &= \mathbf{P}\left(\varrho_x = u_x, \varrho_y = u_y \Big|_{e_x}\right) = \\ &= \begin{cases} \mathbf{P}\left(\varrho_x = u_x \Big|_{u_y=0}\right), & \text{если } u_x \neq 0, u_y = 0 \\ 0, & \text{иначе;} \end{cases} \\ p_y(u_x, u_y) &= \mathbf{P}\left(\varrho_x = u_x, \varrho_y = u_y \Big|_{e_y}\right) = \\ &= \begin{cases} \mathbf{P}\left(\varrho_y = u_y \Big|_{u_x=0}\right), & \text{если } u_x = 0, u_y \neq 0 \\ 0, & \text{иначе;} \end{cases} \end{aligned}$$

$$p_{xy}(u_x, u_y) = \mathbf{P}(\varrho_x = u_x, \varrho_y = u_y |_{e_{xy}}) =$$

$$= \begin{cases} \mathbf{P}(\varrho_x = u_x, \varrho_y = u_y |_{u_x \neq 0, u_y \neq 0}), & \text{если } u_x \neq 0, u_y \neq 0 \\ 0, & \text{иначе,} \end{cases}$$

для любых $u_x, u_y \in (-\infty, \infty)$.

Тогда по теореме о разложении распределения двудольного двумерного случайного вектора распределение двумерного рассматриваемого двудольного случайного вектора имеет вид:

$$\mathbf{P}(\varrho_x = u_x, \varrho_y = u_y) =$$

$$= p(\emptyset)p_\emptyset(u_x, u_y) + p(x)p_x(u_x, u_y) + p(y)p_y(u_x, u_y) + p(xy)p_{xy}(u_x, u_y) =$$

$$= \begin{cases} p(\emptyset), & \text{если } u_x = 0, u_y = 0, \\ p(x)p_x(u_x, u_y), & \text{если } u_x \neq 0, u_y = 0, \\ p(y)p_y(u_x, u_y), & \text{если } u_x = 0, u_y \neq 0, \\ p(xy)p_{xy}(u_x, u_y), & \text{если } u_x \neq 0, u_y \neq 0. \end{cases}$$

Таким образом, вся область определения двумерной вероятности разбивается на четыре области определения условных вероятностей, три из которых несвязны. Область, соответствующая пустому множеству ненулевых случайных величин: \emptyset , связна, области, соответствующие — $\{x\}$ и $\{y\}$, состоят из двух компонент связности, область, соответствующая — $\{x, y\}$, состоит из четырех компонент связности (рис. 2).

$\{x, y\}$ $u_x < 0$ $u_y > 0$	$\{y\}$ $u_x = 0$ $u_y > 0$	$\{x, y\}$ $u_x > 0$ $u_y > 0$
$\{x\}$ $u_x < 0$ $u_y = 0$	\emptyset $u_x = 0$ $u_y = 0$	$\{x\}$ $u_x > 0$ $u_y = 0$
$\{x, y\}$ $u_x < 0$ $u_y < 0$	$\{y\}$ $u_x = 0$ $u_y < 0$	$\{x, y\}$ $u_x > 0$ $u_y < 0$

Рис. 2. Несвязные области определения условных распределений в разложении двудольного двумерного случайного вектора по случайному множественному базису под дуплетом

Тогда совместная функция распределения будет иметь вид

$$F(u_x, u_y) = \mathbf{P}(\varrho_x < u_x, \varrho_y < u_y) = \sum_{v_x < u_x, v_y < u_y} \mathbf{P}(\varrho_x = v_x, \varrho_y = v_y).$$

Несвязность областей определения условных вероятностей требует определенного искусства в задании условных функций распределений для произвольных двудольных случайных векторов.

Теорема. *Количество несвязных областей определения условных плотностей распределения при разложении распределения двудольного случайного вектора по случайно-множественному базису равно $3^{|X|}$.*

Доказательство. Из комбинаторики известна теорема о том, что число различных комбинаций элементов вида (a^1, a^2, \dots, a^r) , где a^l — некоторый элемент l -й группы, состоящей из n^l элементов, равно $n_1 n_2 \dots n_r$.

Рассмотрим $X \in \mathfrak{X}$. Каждую компоненту двудольного вектора можно представить как группу, состоящую из двух элементов: атом и ненулевое значение. Тогда число комбинаций атомов и ненулевых значений двудольного случайного вектора при условии события e_X равно

$$n_1 n_2 \dots n_{|X|} = \underbrace{2 \cdot 2 \dots 2}_{|X|} = 2^{|X|}.$$

Число множеств с мощностью $|X|$ равно $C_{|\mathfrak{X}|}^{|X|}$. Следовательно, количество несвязных областей определения условных распределений

$$\sum_{X \in \mathfrak{X}} 2^{|X|} C_{|\mathfrak{X}|}^{|X|} = 3^{|X|}.$$

Теорема доказана.

Значит, количество несвязных областей определения условных плотностей распределения в разложении двудольного двумерного случайного вектора по случайно-множественному базису под дуплетом равно $3^2 = 9$ (рис. 3). А функция совместного распределения двумерного двудольного случайного вектора, полученная на основе теоремы о разложении распределения двудольного случайного вектора по случайно-множественному базису под дуплетом, имеет вид рис.4. Следует заметить, что при одной и той же количественной надстройке, меняя случайно-множественный базис, будем получать различные функции совместного распределения.

Теорема. *Сумма случайного множества случайных величин $\rho_x, x \in \mathfrak{X}$ равна сумме двудольных случайных величин $\varrho_x, x \in \mathfrak{X}$:*

$$\sum_{x \in K} \rho_x = \sum_{x \in \mathfrak{X}} \varrho_x(p_x),$$

где K — случайное множество ненулевых компонент, а $p_x = \mathbf{P}(x \in K)$ — вероятность покрытия элемента x случайным множеством K .

Доказательство. Каждому $x \in \mathfrak{X}$ можно сопоставить его индикатор — бернуллиевскую случайную величину:

$$\mathbf{1}_x = \begin{cases} 1, & x \in K \\ 0, & \text{иначе,} \end{cases}$$

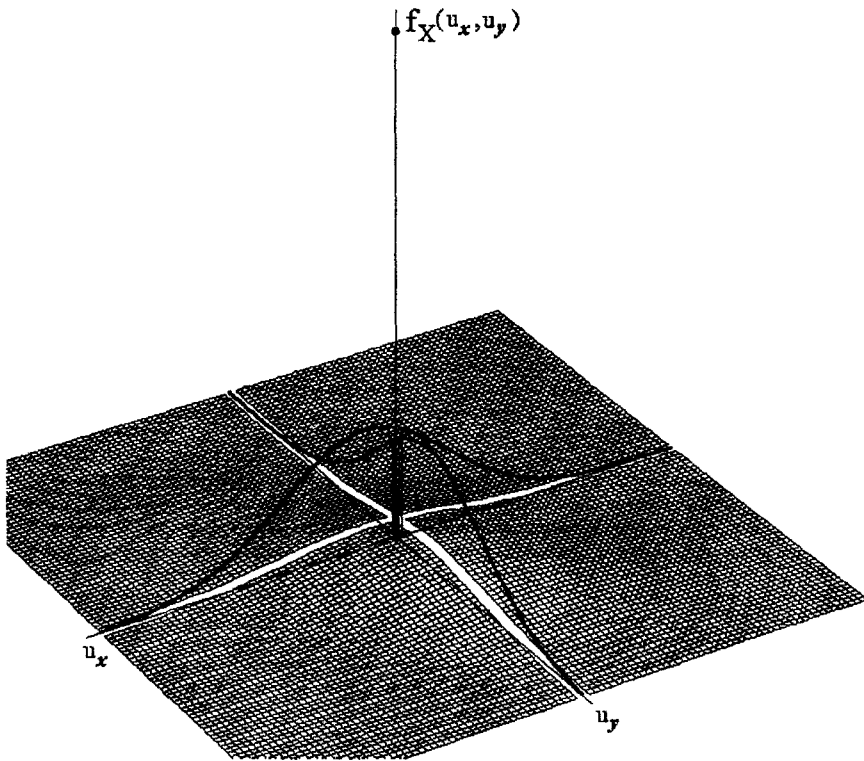


Рис. 3. Несвязные области определения условных плотностей распределения в разложении двудольного двумерного случайного вектора по случайно-множественному базису под дуплетом

с распределением

$$P(\mathbf{1}_x) = \begin{cases} p_x, & x \in K, \\ 1 - p_x, & \text{иначе,} \end{cases}$$

которое определяется вероятностями события x .

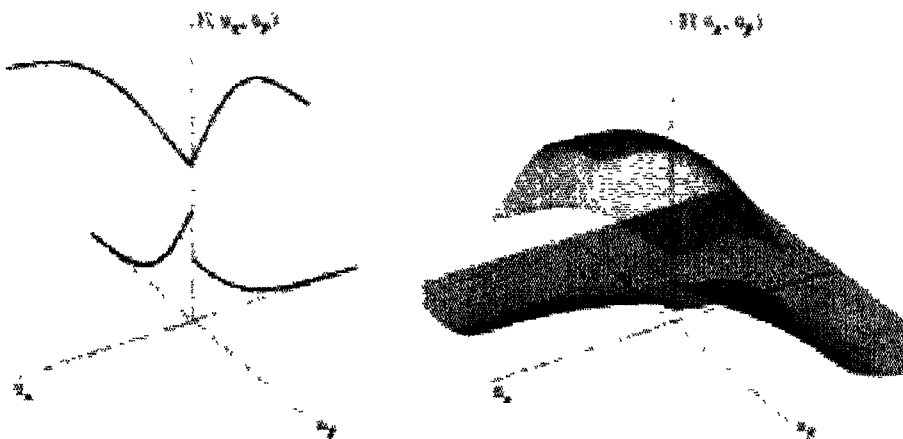


Рис. 4. Функция совместного распределения двумерного двудольного случайного вектора, полученная на основе теоремы о разложении распределения двудольного случайного вектора по случайно-множественному базису: слева – индивидуальные двудольные функции распределения; справа – совместная функция распределения двудольного случайного вектора

Тогда двудольную случайную величину можно записать в виде

$$q_x(p_x) = \rho_x \mathbf{1}_x = \begin{cases} \rho_x, & \text{с вероятностью } p_x, \\ 0, & \text{с вероятностью } 1 - p_x. \end{cases}$$

Тогда

$$\sum_{x \in \mathfrak{X}} q_x(p_x) = \sum_{x \in \mathfrak{X}} \rho_x \mathbf{1}_x = \sum_{x \in K} \rho_x + \sum_{x \in K^c} 0 = \sum_{x \in K} \rho_x.$$

Теорема доказана.

Список литературы

- [1] ВОРОБЬЕВ О.Ю. *Теория случайных событий и ее применения* / О.Ю.Воробьев, Е.Е.Голденюк, Т.В.Куприянова, Д.В.Семенова, А.Ю.Фомин. – Красноярск: ИВМ СО РАН, 2002. – С. 502.
- [2] ГОЛДЕНЮК Е.Е. *О двухуровневой структуре статистических зависимостей и взаимодействий гиббсовских случайных векторов* / Е.Е.Голденюк // Труды I Всероссийской конференции по финансово-актуарной математике и смежным вопросам. – Красноярск: ИВМ СО РАН, 2002. С. 20-34.
- [3] СЕМЕНОВА Д.В. *Теорема о разложении двудольного случайного вектора по случайно-множественному базису* / Д.В.Семенова, Е.Е.Голденюк, О.Ю.Воробьев // Записки ФАМ семинара'2002. – Красноярск: ИВМ СО РАН, 2002. – С. 5-14.