

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

**Факультет информатики, математики и компьютерных наук**

**Программа подготовки бакалавров по направлению  
01.03.02 Прикладная математика и информатика**

*Антонов Илья Витальевич*

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

Проверка условной независимости в трехмерном распределении Бернулли

Научный руководитель

д.ф.-м.н., проф.

П.А. Колданов

Нижний Новгород, 2024

# Содержание

<b>1</b>	<b>Теория проверки условной независимости в трехмерном распределении Бернулли</b>	<b>3</b>
1.1	Условная независимость в трехмерном распределении Бернулли	3
1.2	Частный коэффициент корреляции Пирсона в трехмерном распределении Бернулли . . . . .	5
1.3	Тест на параметр трехмерного распределения Бернулли в экспоненциальной форме . . . . .	10
1.4	Проверка условной независимости по подвыборкам из условных распределений . . . . .	15
	<b>Список использованной литературы</b>	<b>20</b>

# 1 Теория проверки условной независимости в трехмерном распределении Бернулли

## 1.1 Условная независимость в трехмерном распределении Бернулли

Определим трехмерное распределение Бернулли согласно [3], [6].

**Определение 1.1.** Случайный вектор  $(X, Y, Z)^T$  имеет трехмерное распределение Бернулли, если множество его возможных значений:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

и заданы  $P(X = x, Y = y, Z = z) = p_{xyz} \geq 0, \sum_{x=0}^1 \sum_{y=0}^1 \sum_{z=0}^1 p_{xyz} = 1$ .

Приведем определение понятия условной независимости [4].

**Определение 1.2.** Пусть  $(X, Y, Z)^T$  – дискретный случайный вектор. Говорят, что случайные величины  $X$  и  $Y$  условно независимы при условии  $Z$ , и пишут  $X \perp\!\!\!\perp Y \mid Z$ , если:

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)P(Y = y \mid Z = z)$$

при любом  $z$  для которого  $P(Z = z) > 0$ .

Сформулируем и докажем теорему, которая характеризует соотношения между параметрами трехмерного распределения Бернулли при условной независимости.

**Теорема 1.1.** Пусть  $(X, Y, Z)^T$  – случайный вектор, имеющий трехмерное распределение Бернулли, в котором  $P(Z = 0) > 0$ . Случайные величины  $X$  и  $Y$  условно независимы при условии  $Z$  тогда и только тогда, когда  $p_{00z}p_{11z} = p_{01z}p_{10z}$ , где  $z = \overline{0, 1}$ .

*Доказательство.* Пусть  $X \perp\!\!\!\perp Y \mid Z$ . Значит, для любых  $x = \overline{0, 1}$ ,  $y = \overline{0, 1}$  и  $z = \overline{0, 1}$  выполнено условие:

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)P(Y = y \mid Z = z) \quad (1)$$

После домножения (1) на  $P(Z = z)^2$  получаем эквивалентное условие:

$$P(X = x, Y = y, Z = z)P(Z = z) = P(X = x, Z = z)P(Y = y, Z = z) \quad (2)$$

Найдем маргинальное распределение случайной величины  $Z$ :

$$P(Z = z) = \sum_{x=0}^1 \sum_{y=0}^1 p_{xyz} = p_{00z} + p_{01z} + p_{10z} + p_{11z}$$

Найдем маргинальные распределения  $(X, Z)^T$  и  $(Y, Z)^T$ :

$$P(X = x, Z = z) = \sum_{y=0}^1 p_{xyz} = p_{x0z} + p_{x1z}$$

$$P(Y = y, Z = z) = \sum_{x=0}^1 p_{xyz} = p_{0yz} + p_{1yz}$$

Тогда условие (2) перепишем в следующем виде:

$$p_{xyz}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{x0z} + p_{x1z})(p_{0yz} + p_{1yz})$$

Это условие выполняется для всех  $x = \overline{0, 1}$ ,  $y = \overline{0, 1}$ ,  $z = \overline{0, 1}$ . Пусть  $z$  фиксировано. Если  $x = 0$  и  $y = 0$ , то:

$$p_{00z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{00z} + p_{01z})(p_{00z} + p_{10z})$$

$$p_{00z}p_{00z} + p_{00z}p_{01z} + p_{00z}p_{10z} + p_{00z}p_{11z} = p_{00z}p_{00z} + p_{00z}p_{10z} + p_{01z}p_{00z} + p_{01z}p_{10z}$$

$$p_{00z}p_{11z} = p_{01z}p_{10z}$$

Если  $x = 0$  и  $y = 1$ , то:

$$p_{01z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{00z} + p_{01z})(p_{01z} + p_{11z})$$

$$p_{01z}p_{00z} + p_{01z}p_{01z} + p_{01z}p_{10z} + p_{01z}p_{11z} = p_{00z}p_{01z} + p_{00z}p_{11z} + p_{01z}p_{01z} + p_{01z}p_{11z}$$

$$p_{01z}p_{10z} = p_{00z}p_{11z}$$

Если  $x = 1$  и  $y = 0$ , то:

$$p_{10z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{10z} + p_{11z})(p_{00z} + p_{10z})$$

$$p_{10z}p_{00z} + p_{10z}p_{01z} + p_{10z}p_{10z} + p_{10z}p_{11z} = p_{10z}p_{00z} + p_{10z}p_{10z} + p_{11z}p_{00z} + p_{11z}p_{10z}$$

$$p_{10z}p_{01z} = p_{11z}p_{00z}$$

Если  $x = 1$  и  $y = 1$ , то:

$$p_{11z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{10z} + p_{11z})(p_{01z} + p_{11z})$$

$$p_{11z}p_{00z} + p_{11z}p_{01z} + p_{11z}p_{10z} + p_{11z}p_{11z} = p_{10z}p_{01z} + p_{10z}p_{11z} + p_{11z}p_{01z} + p_{11z}p_{11z}$$

$$p_{11z}p_{00z} = p_{10z}p_{01z}$$

Таким образом, из условной независимости  $X$  и  $Y$  при условии  $Z$  следует  $p_{00z}p_{11z} = p_{01z}p_{10z}$ , где  $z = \overline{0, 1}$ .

Доказательство в обратную сторону проводится аналогично.  $\square$

Приведем пример случайного вектора  $(X, Y, Z)^T$  с трехмерным распределением Бернулли, в котором  $X \perp\!\!\!\perp Y \mid Z$ .

**Пример 1.1.** Пусть  $(X, Y, Z)^T$  имеет трехмерное распределение Бернулли с вероятностями  $p_{000} = 0.15$ ,  $p_{001} = 0.1$ ,  $p_{010} = 0.3$ ,  $p_{011} = 0.1$ ,  $p_{100} = 0.05$ ,  $p_{101} = 0.1$ ,  $p_{110} = 0.1$ ,  $p_{111} = 0.1$ . Заметим, что:

$$p_{000}p_{110} = p_{010}p_{100} = 0.015$$

$$p_{001}p_{111} = p_{011}p_{101} = 0.01$$

Значит из теоремы 1.1 следует, что  $X \perp\!\!\!\perp Y \mid Z$ .

## 1.2 Частный коэффициент корреляции Пирсона в трехмерном распределении Бернулли

Согласно [1] в трехмерном нормальном распределении случайные величины  $X$  и  $Y$  условно независимы при условии  $Z$  тогда и только тогда, когда частный коэффициент корреляции Пирсона  $\rho^{XY \cdot Z}$  принимает значение 0. В данном разделе проверим сохраняется ли это свойство в трехмерном распределении Бернулли. Пусть  $(X, Y, Z)^T$  – случайный вектор с трехмерным распределением Бернулли,  $\Sigma$  – ковариационная матрица:

$$\Sigma = \begin{pmatrix} \sigma_{XX} & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{YX} & \sigma_{YY} & \sigma_{YZ} \\ \sigma_{ZX} & \sigma_{ZY} & \sigma_{ZZ} \end{pmatrix}$$

Остатками от  $X$  и  $Y$  при регрессии на  $Z$  называются случайные величины:

$$X' = (X - EX) - \frac{\sigma_{XZ}}{\sigma_{ZZ}}(Z - EZ)$$

$$Y' = (Y - EY) - \frac{\sigma_{YZ}}{\sigma_{ZZ}}(Z - EZ)$$

Согласно работе [2] частным коэффициентом корреляции Пирсона называется:

$$\rho^{XY \cdot Z} = \frac{E(X'Y')}{\sqrt{E(X')^2 E(Y')^2}}$$

Для частного коэффициента корреляции Пирсона установим явную формулу.

**Лемма 1.1.**

$$\rho^{XY \cdot Z} = \frac{\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ}}{\sqrt{\sigma_{XX}\sigma_{ZZ} - \sigma_{XZ}^2}\sqrt{\sigma_{YY}\sigma_{ZZ} - \sigma_{YZ}^2}}$$

*Доказательство.*

$$\begin{aligned} E(X'Y') &= E\left[\left((X - EX) - \frac{\sigma_{XZ}}{\sigma_{ZZ}}(Z - EZ)\right)\left((Y - EY) - \frac{\sigma_{YZ}}{\sigma_{ZZ}}(Z - EZ)\right)\right] = \\ &= \sigma_{XY} - \frac{\sigma_{YZ}}{\sigma_{ZZ}}\sigma_{XZ} - \frac{\sigma_{XZ}}{\sigma_{ZZ}}\sigma_{YZ} + \frac{\sigma_{XZ}\sigma_{YZ}}{\sigma_{ZZ}\sigma_{ZZ}}\sigma_{ZZ} = \sigma_{XY} - \frac{\sigma_{XZ}\sigma_{YZ}}{\sigma_{ZZ}} = \\ &= \frac{\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ}}{\sigma_{ZZ}} \end{aligned}$$

$$\begin{aligned} \rho^{XY \cdot Z} &= \frac{\frac{\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ}}{\sigma_{ZZ}}}{\sqrt{\frac{\sigma_{XX}\sigma_{ZZ} - \sigma_{XZ}^2}{\sigma_{ZZ}}}\sqrt{\frac{\sigma_{YY}\sigma_{ZZ} - \sigma_{YZ}^2}{\sigma_{ZZ}}}} = \\ &= \frac{\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ}}{\sqrt{\sigma_{XX}\sigma_{ZZ} - \sigma_{XZ}^2}\sqrt{\sigma_{YY}\sigma_{ZZ} - \sigma_{YZ}^2}} \end{aligned}$$

□

Для дальнейших рассуждений удобно принять следующие обозначения:

$$p_{x**} = P(X = x), \quad p_{*y*} = P(Y = y), \quad p_{**z} = P(Z = z)$$

$$p_{xy*} = P(X = x, Y = y), \quad p_{x*z} = P(X = x, Z = z), \quad p_{*yz} = P(Y = y, Z = z)$$

Путем алгебраических преобразований, установим соотношение для числителя частного коэффициента корреляции Пирсона в трехмерном распределении Бернулли.

**Лемма 1.2.**

$$\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ} = p_{**0}(p_{001}p_{111} - p_{011}p_{101}) + p_{**1}(p_{000}p_{110} - p_{010}p_{100})$$

*Доказательство.* Легко проверить, что  $\sigma_{XX} = p_{1**}(1 - p_{1**})$ . Найдем соотношение для  $\sigma_{XY}$ . Воспользуемся формулой  $\sigma_{XY} = E(XY) - E(X)E(Y)$ .

$$E(XY) = 1 \cdot p_{11*} + 0 \cdot (p_{00*} + p_{01*} + p_{10*}) = p_{11*}$$

$$EX = 1 \cdot p_{1**} + 0 \cdot p_{0**} = p_{1**}$$

$$EY = 1 \cdot p_{*1*} + 0 \cdot p_{*0*} = p_{*1*}$$

Таким образом,  $\sigma_{XY} = p_{11*} - p_{1**}p_{*1*}$ . Аналогично,  $\sigma_{XZ} = p_{1*1} - p_{1**}p_{**1}$  и  $\sigma_{YZ} = p_{*11} - p_{*1*}p_{**1}$ . Непосредственно преобразуем выражение:

$$\begin{aligned} & \sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ} = \\ &= (p_{11*} - p_{1**}p_{*1*})p_{**1}(1 - p_{**1}) - (p_{1*1} - p_{1**}p_{**1})(p_{*11} - p_{*1*}p_{**1}) = \\ &= p_{11*}p_{**1} - p_{11*}p_{**1}p_{**1} - p_{1**}p_{*1*}p_{**1} + p_{1**}p_{*1*}p_{**1}p_{**1} - \\ & - p_{1*1}p_{*11} + p_{1*1}p_{*1*}p_{**1} + p_{1**}p_{**1}p_{*11} - p_{1**}p_{**1}p_{*1*}p_{**1} = \\ &= p_{111}p_{**1} + p_{110}p_{**1} - p_{11*}p_{**1}p_{**1} - p_{1**}p_{*1*}p_{**1} - \\ & - p_{1*1}p_{*11} + p_{1*1}p_{*1*}p_{**1} + p_{1**}p_{**1}p_{*11} = \\ &= (p_{111}p_{**1} - p_{1*1}p_{*11}) + p_{**1}(p_{110} - p_{11*}p_{**1} - p_{1**}p_{*1*} + p_{1*1}p_{*1*} + p_{1**}p_{*11}) \quad (3) \end{aligned}$$

Заметим, что:

1.

$$\begin{aligned} p_{110} - p_{11*}p_{**1} &= p_{110} - p_{110}p_{**1} - p_{111}p_{**1} = p_{110}(1 - p_{**1}) - p_{111}p_{**1} = \\ &= p_{110}p_{**0} - p_{111}p_{**1} \end{aligned}$$

2.

$$\begin{aligned} & -p_{1**}p_{*1*} + p_{1*1}p_{*1*} + p_{1**}p_{*11} = \\ &= -(p_{1*0} + p_{1*1})(p_{*10} + p_{*11}) + p_{1*1}(p_{*10} + p_{*11}) + (p_{1*0} + p_{1*1})p_{*11} = \\ &= -p_{1*0}p_{*10} - p_{1*0}p_{*11} - p_{1*1}p_{*10} - p_{1*1}p_{*11} + \\ & + p_{1*1}p_{*10} + p_{1*1}p_{*11} + p_{1*0}p_{*11} + p_{1*1}p_{*11} = \\ &= -p_{1*0}p_{*10} + p_{1*1}p_{*11} \end{aligned}$$

Учитывая вышеприведенные соотношения, запишем выражение (3).

$$\begin{aligned}
& (p_{111}p_{**1} - p_{1*1}p_{*11}) + p_{**1}((p_{110}p_{**0} - p_{1*0}p_{*10}) - (p_{111}p_{**1} - p_{1*1}p_{*11})) = \\
& = (p_{111}p_{**1} - p_{1*1}p_{*11}) + p_{**1}(p_{110}p_{**0} - p_{1*0}p_{*10}) - p_{**1}(p_{111}p_{**1} - p_{1*1}p_{*11}) = \\
& = (1 - p_{**1})(p_{111}p_{**1} - p_{1*1}p_{*11}) + p_{**1}(p_{110}p_{**0} - p_{1*0}p_{*10}) = \\
& = p_{**0}(p_{111}p_{**1} - p_{1*1}p_{*11}) + p_{**1}(p_{110}p_{**0} - p_{1*0}p_{*10})
\end{aligned}$$

Также заметим, что:

1.

$$\begin{aligned}
p_{111}p_{**1} - p_{1*1}p_{*11} & = p_{111}(p_{001} + p_{011} + p_{101} + p_{111}) - (p_{101} + p_{111})(p_{011} + p_{111}) = \\
& = p_{111}p_{001} + p_{111}p_{011} + p_{111}p_{101} + p_{111}p_{111} - \\
& - p_{101}p_{011} - p_{101}p_{111} - p_{111}p_{011} - p_{111}p_{111} = \\
& = p_{001}p_{111} - p_{011}p_{101}
\end{aligned}$$

2.

$$\begin{aligned}
p_{110}p_{**0} - p_{1*0}p_{*10} & = p_{110}(p_{000} + p_{010} + p_{100} + p_{110}) - (p_{100} + p_{110})(p_{010} + p_{110}) = \\
& = p_{110}p_{000} + p_{110}p_{010} + p_{110}p_{100} + p_{110}p_{110} - \\
& - p_{100}p_{010} - p_{100}p_{110} - p_{110}p_{010} - p_{110}p_{110} = \\
& = p_{000}p_{110} - p_{010}p_{100}
\end{aligned}$$

Таким образом:

$$\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ} = p_{**0}(p_{001}p_{111} - p_{011}p_{101}) + p_{**1}(p_{000}p_{110} - p_{010}p_{100})$$

□

Вышеприведенное соотношение позволяет доказать следующую теорему.

**Теорема 1.2.** Пусть  $X$  и  $Y$  условно независимы при условии  $Z$ . Тогда  $\rho^{XY \cdot Z} = 0$ .



*Доказательство.* Пусть  $X$  и  $Y$  условно независимы при условии  $Z$ . Тогда по теореме 1.1:  $p_{000}p_{110} = p_{010}p_{100}$  и  $p_{001}p_{111} = p_{011}p_{101}$ . Используя эти соотношения в числителе частного коэффициента корреляции Пирсона имеем:

$$p_{**0}(p_{001}p_{111} - p_{011}p_{101}) + p_{**1}(p_{000}p_{110} - p_{010}p_{100}) = 0$$

Следовательно,  $\rho^{XY \cdot Z} = 0$ . □

Таким образом, ноль в частном коэффициенте корреляции Пирсона является необходимым условием условной независимости. Однако, не является достаточным условием, так как в обратную сторону теорема 1.2 неверна. Легко построить контрпример при  $p_{**0} = 0$ . Далее покажем контрпример при  $p_{**0} \neq 0$ .

**Пример 1.2.** Пусть  $p_{000} = 0.15$ ,  $p_{001} = 0.1$ ,  $p_{010} = 0.1$ ,  $p_{011} = 0.15$ ,  $p_{100} = 0.1$ ,  $p_{101} = 0.15$ ,  $p_{110} = 0.15$ ,  $p_{111} = 0.1$ . Тогда  $p_{**0} = 0.5$ ,  $p_{**1} = 0.5$  и  $M_{21} = p_{**1}(p_{000}p_{110} - p_{010}p_{100}) + p_{**0}(p_{001}p_{111} - p_{011}p_{101}) = 0.5 \cdot (0.15 \cdot 0.15 - 0.1 \cdot 0.1) + 0.5 \cdot (0.1 \cdot 0.1 - 0.15 \cdot 0.15) = 0$ .

Однако, случайные величины  $X$  и  $Y$  условно зависимы при условии  $Z$  поскольку:

$$p_{000}p_{110} - p_{010}p_{100} = 0.15 \cdot 0.15 - 0.1 \cdot 0.1 = 0.0125 \neq 0$$

$$p_{001}p_{111} - p_{011}p_{101} = 0.1 \cdot 0.1 - 0.15 \cdot 0.15 = -0.0125 \neq 0$$

Приведем альтернативную формулу, с помощью которой удобно вычислять частный коэффициент корреляции Пирсона.

**Лемма 1.3.**

$$\rho^{XY \cdot Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1 - \rho_{XZ}^2}\sqrt{1 - \rho_{YZ}^2}}$$

где  $\rho_{XY}, \rho_{XZ}, \rho_{YZ}$  — коэффициенты корреляции Пирсона между случайными величинами  $X$  и  $Y$ ,  $X$  и  $Z$ ,  $Y$  и  $Z$  соответственно.

*Доказательство.*

$$\rho^{XY \cdot Z} = \frac{\sigma_{XY}\sigma_{ZZ} - \sigma_{XZ}\sigma_{YZ}}{\sqrt{\sigma_{XX}\sigma_{ZZ} - \sigma_{XZ}^2}\sqrt{\sigma_{YY}\sigma_{ZZ} - \sigma_{YZ}^2}}$$

$$\begin{aligned}
& \frac{\sigma_{ZZ}\sqrt{\sigma_{XX}\sigma_{YY}}\left(\frac{\sigma_{XY}}{\sqrt{\sigma_{XX}\sigma_{YY}}} - \frac{\sigma_{XZ}}{\sqrt{\sigma_{XX}\sigma_{ZZ}}}\frac{\sigma_{YZ}}{\sqrt{\sigma_{YY}\sigma_{ZZ}}}\right)}{\sqrt{\sigma_{XX}\sigma_{ZZ} - \sigma_{XZ}^2}\sqrt{\sigma_{YY}\sigma_{ZZ} - \sigma_{YZ}^2}} = \\
& = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1 - \frac{\sigma_{XZ}^2}{\sigma_{XX}\sigma_{ZZ}}}\sqrt{1 - \frac{\sigma_{YZ}^2}{\sigma_{YY}\sigma_{ZZ}}}} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1 - \rho_{XZ}^2}\sqrt{1 - \rho_{YZ}^2}}
\end{aligned}$$

□

Классически для оценки частного коэффициента корреляции Пирсона используют выборочный частный коэффициент корреляции Пирсона:

$$r^{XY \cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{1 - r_{XZ}^2}\sqrt{1 - r_{YZ}^2}}$$

где  $r_{XY}$ ,  $r_{XZ}$ ,  $r_{YZ}$  – выборочные коэффициенты корреляции Пирсона между случайными величинами  $X$  и  $Y$ ,  $X$  и  $Z$ ,  $Y$  и  $Z$  соответственно.

Для нормального распределения известно [1], что при истинности гипотезы  $\rho^{XY \cdot Z} = 0$  статистика

$$T = \sqrt{n-3} \frac{r^{XY \cdot Z}}{\sqrt{1 - (r^{XY \cdot Z})^2}}$$

имеет распределение Стьюдента с  $n-3$  степенями свободы. Тогда тест уровня  $\alpha$  проверки гипотезы  $H : \rho^{XY \cdot Z} = 0$  против альтернативы  $K : \rho^{XY \cdot Z} \neq 0$  имеет вид:

$$\varphi^{\text{Partial}}(t) = \begin{cases} 1, & t < C_1 \text{ или } t > C_2 \\ 0, & C_1 \leq t \leq C_2 \end{cases}$$

где константы  $C_1$  и  $C_2$ , удовлетворяют уравнениям  $P(T < C_1) = \alpha/2$  и  $P(T > C_2) = 1 - \alpha/2$ .

В настоящей работе с помощью численных экспериментов будет проверено контролирует ли тест  $\varphi^{\text{Partial}}$  вероятность ошибки первого рода в трехмерном распределении Бернулли.

### 1.3 Тест на параметр трехмерного распределения Бернулли в экспоненциальной форме

Следующая лемма характеризует вид трехмерного распределения Бернулли в экспоненциальной форме.

**Лемма 1.4.**

$$\begin{aligned}
P(X = x, Y = y, Z = z) = p_{000} \exp \left\{ xyz \ln \left( \frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}} \right) + \right. \\
\left. + x \ln \left( \frac{p_{100}}{p_{000}} \right) + y \ln \left( \frac{p_{010}}{p_{000}} \right) + z \ln \left( \frac{p_{001}}{p_{000}} \right) + \right. \\
\left. + xy \ln \left( \frac{p_{000}p_{110}}{p_{010}p_{100}} \right) + xz \ln \left( \frac{p_{000}p_{101}}{p_{001}p_{100}} \right) + yz \ln \left( \frac{p_{000}p_{011}}{p_{001}p_{010}} \right) \right\}
\end{aligned}$$

*Доказательство.*

$$\begin{aligned}
P(X = x, Y = y, Z = z) &= p_{000}^{(1-x)(1-y)(1-z)} \dots p_{111}^{xyz} = \\
&= \exp \left\{ (1-x)(1-y)(1-z) \ln p_{000} + (1-x)(1-y)z \ln p_{001} + \right. \\
&+ (1-x)y(1-z) \ln p_{010} + (1-x)yz \ln p_{011} + x(1-y)(1-z) \ln p_{100} + \\
&\left. + x(1-y)z \ln p_{101} + xy(1-z) \ln p_{110} + xyz \ln p_{111} \right\} = \\
&= \exp \left\{ (1-y-x+xy-z+yz+xz-xyz) \ln p_{000} + \right. \\
&+ (z-yz-xz+xyz) \ln p_{001} + (y-yz-xy+xyz) \ln p_{010} + \\
&+ (yz-xyz) \ln p_{011} + (x-xz-xy+xyz) \ln p_{100} + \\
&\left. + (xz-xyz) \ln p_{101} + (xy-xyz) \ln p_{110} + xyz \ln p_{111} \right\} = \\
&= p_{000} \exp \left\{ xyz \ln \left( \frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}} \right) + \right. \\
&+ x \ln \left( \frac{p_{100}}{p_{000}} \right) + y \ln \left( \frac{p_{010}}{p_{000}} \right) + z \ln \left( \frac{p_{001}}{p_{000}} \right) + \\
&\left. + xy \ln \left( \frac{p_{000}p_{110}}{p_{010}p_{100}} \right) + xz \ln \left( \frac{p_{000}p_{101}}{p_{001}p_{100}} \right) + yz \ln \left( \frac{p_{000}p_{011}}{p_{001}p_{010}} \right) \right\}
\end{aligned}$$

□

Среди параметров, стоящих при статистиках можно выделить параметр, связанный с условной независимостью.

**Теорема 1.3.** Пусть  $\theta = \ln \left( \frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}} \right)$ . Если выполнено одно из условий:

- $X \perp\!\!\!\perp Y \mid Z$
- $X \perp\!\!\!\perp Z \mid Y$
- $Y \perp\!\!\!\perp Z \mid X$

то параметр  $\theta$  принимает значение 0.

*Доказательство.* Результаты теоремы 1.1 можно обобщить следующим образом:

$$X \perp\!\!\!\perp Z \mid Y \Leftrightarrow p_{000}p_{101} = p_{001}p_{100} \text{ и } p_{010}p_{111} = p_{011}p_{110}$$

$$Y \perp\!\!\!\perp Z \mid X \Leftrightarrow p_{000}p_{011} = p_{001}p_{010} \text{ и } p_{100}p_{111} = p_{101}p_{110}$$

1. Пусть  $X \perp\!\!\!\perp Y \mid Z$ , тогда по теореме 1.1 выполнено:  $p_{000}p_{110} = p_{010}p_{100}$  и  $p_{001}p_{111} = p_{011}p_{101}$ . Отсюда следует, что  $\theta = \ln(1) = 0$ .
2. Пусть  $X \perp\!\!\!\perp Z \mid Y$ , тогда из вышеприведенных соображений  $p_{000}p_{101} = p_{001}p_{100}$  и  $p_{010}p_{111} = p_{011}p_{110}$ . Отсюда следует, что  $\theta = \ln(1) = 0$ .
3. Пусть  $Y \perp\!\!\!\perp Z \mid X$ , тогда из вышеприведенных соображений  $p_{000}p_{011} = p_{001}p_{010}$  и  $p_{100}p_{111} = p_{101}p_{110}$ . Отсюда следует, что  $\theta = \ln(1) = 0$ .

□

Таким образом, параметр  $\theta = \ln \left( \frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}} \right)$  отражает наличие условно независимой пары случайных величин в трехмерном распределении Бернулли. Для проверки гипотезы о равенстве параметра  $\theta$  нулю можно использовать теорию РНМН тестов [5]. Пусть

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \\ Z_n \end{pmatrix}$$

повторная выборка из распределения случайного вектора  $(X, Y, Z)^T$ . Из леммы 1.4 непосредственно следует, что совместное распределение повторной выборки имеет вид:

$$P(X_1 = x_1, Y_1 = y_1, Z_1 = z_1, \dots, X_n = x_n, Y_n = y_n, Z_n = z_n) =$$

$$\begin{aligned}
&= \prod_{i=1}^n P(X_i = x_i, Y_i = y_i, Z_i = z_i) = \\
&= (p_{000})^n \exp \left\{ \ln \left( \frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}} \right) \sum_{i=1}^n x_i y_i z_i + \right. \\
&\quad + \ln \left( \frac{p_{100}}{p_{000}} \right) \sum_{i=1}^n x_i + \ln \left( \frac{p_{010}}{p_{000}} \right) \sum_{i=1}^n y_i + \ln \left( \frac{p_{001}}{p_{000}} \right) \sum_{i=1}^n z_i + \\
&\quad \left. + \ln \left( \frac{p_{000}p_{110}}{p_{010}p_{100}} \right) \sum_{i=1}^n x_i y_i + \ln \left( \frac{p_{000}p_{101}}{p_{001}p_{100}} \right) \sum_{i=1}^n x_i z_i + \ln \left( \frac{p_{000}p_{011}}{p_{001}p_{010}} \right) \sum_{i=1}^n y_i z_i \right\}
\end{aligned}$$

Согласно [5] РНМН тест проверки гипотезы  $H : \theta = \theta_0$  против альтернативы  $K : \theta \neq \theta_0$ , где  $\theta_0 = 0$ , имеет вид:

$$\varphi^{\theta - \text{UMPU}}(u, t) = \begin{cases} 1, & u < C_1(t) \text{ или } u > C_2(t) \\ \gamma_i, & u = C_i(t), \quad i = 1, 2 \\ 0, & C_1(t) < u < C_2(t) \end{cases}$$

где константы  $C_i$  и  $\gamma_i$  определяются из системы уравнений:

$$\begin{cases} E_{\theta_0}[\varphi^{\theta - \text{UMPU}}(U, T) \mid T = t] = \alpha \\ E_{\theta_0}[U \varphi^{\theta - \text{UMPU}}(U, T) \mid T = t] = \alpha E_{\theta_0}[U \mid T = t] \end{cases}$$

а статистиками являются:

$$U = \sum_{i=1}^n X_i Y_i Z_i, T_1 = \sum_{i=1}^n X_i Y_i, T_2 = \sum_{i=1}^n X_i Z_i, T_3 = \sum_{i=1}^n Y_i Z_i,$$

$$T_4 = \sum_{i=1}^n X_i, T_5 = \sum_{i=1}^n Y_i, T_6 = \sum_{i=1}^n Z_i$$

и  $T = (T_1, \dots, T_6)$  и  $t = (t_1, \dots, t_6)$ . Положим  $k_1(u) = u$ ,  $k_2(u) = t_1 - u$ ,  $k_3(u) = t_2 - u$ ,  $k_4(u) = t_3 - u$ ,  $k_5(u) = t_4 - t_1 - t_2 + u$ ,  $k_6(u) = t_5 - t_1 - t_3 + u$ ,  $k_7(u) = t_6 - t_2 - t_3 + u$ ,  $k_8(u) = n - u + t_1 + t_2 + t_3 - t_4 - t_5 - t_6$ . Приведем лемму, характеризующую совместное распределение статистик  $(U, T_1, T_2, T_3, T_4, T_5, T_6)$ .

**Лемма 1.5.**

$$P(U = u, T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) =$$

$$\begin{aligned}
&= \frac{n!}{\prod_{i=1}^8 k_i(u)!} \left( \frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \right)^u \left( \frac{p_{000}p_{110}}{p_{010}p_{100}} \right)^{t_1} \left( \frac{p_{000}p_{101}}{p_{001}p_{100}} \right)^{t_2} \left( \frac{p_{000}p_{011}}{p_{001}p_{010}} \right)^{t_3} \times \\
&\quad \times \left( \frac{p_{100}}{p_{000}} \right)^{t_4} \left( \frac{p_{010}}{p_{000}} \right)^{t_5} \left( \frac{p_{001}}{p_{000}} \right)^{t_6} p_{000}^n
\end{aligned}$$

*Доказательство.*

$$\begin{aligned}
&P(U = u, T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \\
&= P\left(\sum_{i=1}^n X_i Y_i Z_i = u, \sum_{i=1}^n X_i Y_i = t_1, \sum_{i=1}^n X_i Z_i = t_2, \sum_{i=1}^n Y_i Z_i = t_3, \right. \\
&\quad \left. \sum_{i=1}^n X_i = t_4, \sum_{i=1}^n Y_i = t_5, \sum_{i=1}^n Z_i = t_6\right) = \\
&= P\left(\sum_{i=1}^n X_i Y_i Z_i = u, \sum_{i=1}^n X_i Y_i (1 - Z_i) = t_1 - u, \sum_{i=1}^n X_i (1 - Y_i) Z_i = t_2 - u, \right. \\
&\quad \sum_{i=1}^n (1 - X_i) Y_i Z_i = t_3 - u, \sum_{i=1}^n X_i (1 - Y_i) (1 - Z_i) = t_4 - t_1 - t_2 + u, \\
&\quad \sum_{i=1}^n (1 - X_i) Y_i (1 - Z_i) = t_5 - t_1 - t_3 + u, \sum_{i=1}^n (1 - X_i) (1 - Y_i) Z_i = t_6 - t_2 - t_3 + u, \\
&\quad \left. \sum_{i=1}^n (1 - X_i) (1 - Y_i) (1 - Z_i) = n - u + t_1 + t_2 + t_3 - t_4 - t_5 - t_6\right) = \\
&= \frac{n!}{\prod_{i=1}^8 k_i(u)!} p_{111}^u p_{110}^{t_1-u} p_{101}^{t_2-u} p_{011}^{t_3-u} p_{100}^{t_4-t_1-t_2+u} p_{010}^{t_5-t_1-t_3+u} p_{001}^{t_6-t_2-t_3+u} \times \\
&\quad \times p_{000}^{n-u+t_1+t_2+t_3-t_4-t_5-t_6} = \\
&= \frac{n!}{\prod_{i=1}^8 k_i(u)!} \left( \frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \right)^u \left( \frac{p_{000}p_{110}}{p_{010}p_{100}} \right)^{t_1} \left( \frac{p_{000}p_{101}}{p_{001}p_{100}} \right)^{t_2} \left( \frac{p_{000}p_{011}}{p_{001}p_{010}} \right)^{t_3} \times \\
&\quad \times \left( \frac{p_{100}}{p_{000}} \right)^{t_4} \left( \frac{p_{010}}{p_{000}} \right)^{t_5} \left( \frac{p_{001}}{p_{000}} \right)^{t_6} p_{000}^n
\end{aligned}$$

□

Явно найдем распределение  $U = u$  при условии  $T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6$ .

**Лемма 1.6.**

$$P_{\theta_0}(U = u \mid T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \frac{\frac{1}{\prod_{i=1}^8 k_i(u)!}}{\sum_{s \in \mathcal{D}} \frac{1}{\prod_{i=1}^8 k_i(s)!}}$$

где  $\mathcal{D} = \{x \in \mathbb{Z} : 0 \leq k_i(s) \leq n \text{ для всех } i = 1 \dots, 8\}$ .

*Доказательство.* Найдем маргинальное распределение вектора  $(T_1, \dots, T_6)^T$ :

$$\begin{aligned} P(T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \\ = \sum_{s \in \mathcal{D}} P(U = s, T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) \end{aligned}$$

Тогда условное распределение статистики  $U$  при условии  $T_1 = t_1, \dots, T_6 = t_6$  можно записать в виде:

$$\begin{aligned} P(U = u \mid T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \\ = \frac{\frac{n!}{\prod_{i=1}^8 k_i(u)!} \left( \frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \right)^u}{\sum_{s \in \mathcal{D}} \frac{n!}{\prod_{i=1}^8 k_i(s)!} \left( \frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \right)^s} \end{aligned}$$

При истинности гипотезы  $\theta = \theta_0 = 0$  параметр  $\frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} = 1$ .

$$P_{\theta_0}(U = u \mid T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \frac{\frac{1}{\prod_{i=1}^8 k_i(u)!}}{\sum_{s \in \mathcal{D}} \frac{1}{\prod_{i=1}^8 k_i(s)!}}$$

□

## 1.4 Проверка условной независимости по подвыборкам из условных распределений

Пусть

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \\ Z_n \end{pmatrix}$$

повторная выборка из распределения случайного вектора  $(X, Y, Z)^T$ . Обозначим  $\mathbf{Z} = (Z_1, \dots, Z_n)$  и  $\mathbf{z} = (z_1, \dots, z_n)$ . Покажем, что в условном распределении при  $\mathbf{Z} = \mathbf{z}$  наблюдения:

$$\Xi = \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$$

являются независимыми.

**Лемма 1.7.**

$$\begin{aligned} P(X_1 = x_1, Y_1 = y_1, \dots, X_n = x_n, Y_n = y_n \mid \mathbf{Z} = \mathbf{z}) &= \\ &= \prod_{i=1}^n P(X_i = x_i, Y_i = y_i \mid \mathbf{Z} = \mathbf{z}) \end{aligned}$$

*Доказательство.* С одной стороны:

$$\begin{aligned} P(X_1 = x_1, Y_1 = y_1, Z_1 = z_1, \dots, X_n = x_n, Y_n = y_n, Z_n = z_n) &= \\ &= P(\mathbf{Z} = \mathbf{z})P(X_1 = x_1, Y_1 = y_1, \dots, X_n = x_n, Y_n = y_n \mid \mathbf{Z} = \mathbf{z}) \end{aligned}$$

С другой стороны:

$$\begin{aligned} P(X_1 = x_1, Y_1 = y_1, Z_1 = z_1, \dots, X_n = x_n, Y_n = y_n, Z_n = z_n) &= \\ &= \prod_{i=1}^n P(X_i = x_i, Y_i = y_i, Z_i = z_i) = \\ &= \prod_{i=1}^n P(X_i = x_i, Y_i = y_i \mid Z_i = z_i)P(Z_i = z_i) = \\ &= P(\mathbf{Z} = \mathbf{z}) \prod_{i=1}^n P(X_i = x_i, Y_i = y_i \mid \mathbf{Z} = \mathbf{z}) \end{aligned}$$

□

Пусть также  $\mathbf{Z} = \mathbf{z}$  фиксированы. Разобьем выборку  $\Xi$  на две подвыборки  $\Xi_0$  и  $\Xi_1$ , такие что:

$$\Xi_0 = \begin{pmatrix} X_{i_1} \\ Y_{i_1} \end{pmatrix}, \begin{pmatrix} X_{i_2} \\ Y_{i_2} \end{pmatrix}, \dots, \begin{pmatrix} X_{i_{n_0}} \\ Y_{i_{n_0}} \end{pmatrix}$$



где  $z_{i_k} = 0$  для всех  $i_k$  при  $k = \overline{1, n_0}$  и

$$\Xi_1 = \begin{pmatrix} X_{j_1} \\ Y_{j_1} \end{pmatrix}, \begin{pmatrix} X_{j_2} \\ Y_{j_2} \end{pmatrix}, \dots, \begin{pmatrix} X_{j_{n_1}} \\ Y_{j_{n_1}} \end{pmatrix}$$

где  $z_{j_k} = 1$  для всех  $j_k$  при  $k = \overline{1, n_1}$ . Сформулируем следующую теорему.

**Теорема 1.4.** Пусть  $\mathbf{Z} = \mathbf{z}$  – фиксированы. Тогда выборка  $\Xi_0$  является повторной выборкой из распределения  $(X, Y)^T$  при условии  $Z = 0$ .

*Доказательство.* По лемме 1.7:

$$\begin{aligned} P(X_1 = x_1, Y_1 = y_1, \dots, X_n = x_n, Y_n = y_n \mid \mathbf{Z} = \mathbf{z}) = \\ = \prod_{i=1}^n P(X_i = x_i, Y_i = y_i \mid \mathbf{Z} = \mathbf{z}) \end{aligned}$$

Просуммировав обе части вышепредставленного равенства по всем возможным значениям  $x_{j_1}, y_{j_1}, \dots, x_{j_{n_1}}, y_{j_{n_1}}$  получаем:

$$\begin{aligned} P(X_{i_1} = x_{i_1}, Y_{i_1} = y_{i_1}, \dots, X_{i_{n_0}} = x_{i_{n_0}}, Y_{i_{n_0}} = y_{i_{n_0}} \mid \mathbf{Z} = \mathbf{z}) = \\ = \prod_{k=1}^{n_0} P(X_{i_{n_k}} = x_{i_{n_k}}, Y_{i_{n_k}} = y_{i_{n_k}} \mid \mathbf{Z} = \mathbf{z}) \end{aligned}$$

Значит

$$\Xi_0 = \begin{pmatrix} X_{i_1} \\ Y_{i_1} \end{pmatrix}, \begin{pmatrix} X_{i_2} \\ Y_{i_2} \end{pmatrix}, \dots, \begin{pmatrix} X_{i_{n_0}} \\ Y_{i_{n_0}} \end{pmatrix}$$

независимые наблюдения при условии  $\mathbf{Z} = \mathbf{z}$ .

Покажем, что  $\begin{pmatrix} X_{i_k} \\ Y_{i_k} \end{pmatrix}$  при условии  $\mathbf{Z} = \mathbf{z}$  распределен также как и  $\begin{pmatrix} X \\ Y \end{pmatrix}$  при условии  $Z = 0$ .

$$\begin{aligned} P(X_{i_{n_k}} = x_{i_{n_k}}, Y_{i_{n_k}} = y_{i_{n_k}} \mid Z_1 = z_1, \dots, Z_n = z_n) = \\ = P(X_{i_{n_k}} = x_{i_{n_k}}, Y_{i_{n_k}} = y_{i_{n_k}} \mid Z_{i_{n_k}} = z_{i_{n_k}}) = \\ = P(X = x_{i_{n_k}}, Y = y_{i_{n_k}} \mid Z_{i_{n_k}} = 0) \end{aligned}$$

□

Аналогично показывается, что выборка  $\Xi_1$  является повторной выборкой из распределения  $(X, Y)^T$  при условии  $Z = 1$ .

**Теорема 1.5.** Пусть  $\Xi_0$  и  $\Xi_1$  – подвыборки, полученные разбиением случайной выборки  $\Xi$ . Пусть  $\varphi_0$  – рандомизированный тест проверки произвольной гипотезы  $h_0$  по подвыборке  $\Xi_0$ , причем  $\varphi_0$  тест уровня при любом размере подвыборки  $\Xi_0$ . Пусть  $\varphi_1$  – рандомизированный тест проверки произвольной гипотезы  $h_1$  уровня  $\alpha_1$  по подвыборке  $\Xi_1$ , причем  $\varphi_1$  тест уровня  $\alpha_1$  при любом размере подвыборки  $\Xi_1$ . Обозначим

$$A_0 = \{\text{отвергнуть гипотезу } h_0 \text{ рандомизированным тестом } \varphi_0\}$$

$$A_1 = \{\text{отвергнуть гипотезу } h_1 \text{ рандомизированным тестом } \varphi_1\}$$

Тогда  $P_{h_0 \cap h_1}(A_0 \cap A_1) = P_{h_0 \cap h_1}(A_0)P_{h_0 \cap h_1}(A_1)$ .

*Доказательство.* Пусть  $\mathbf{Z} = \mathbf{z}$  фиксированы, тогда статистиками тестов  $\varphi_0 = \varphi_0(t)$  и  $\varphi_1 = \varphi_1(t)$  являются  $T_0 = T_0(X_{i_1}, Y_{i_1}, \dots, X_{i_{n_0}}, Y_{i_{n_0}})$ ,  $T_1 = T_1(X_{j_1}, Y_{j_1}, \dots, X_{j_{n_1}}, Y_{j_{n_1}})$  соответственно. Отметим, что  $T_0$  и  $T_1$  – независимы при условии  $\mathbf{Z} = \mathbf{z}$ , поскольку наблюдения

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$$

независимы при условии  $\mathbf{Z} = \mathbf{z}$  и аргументы статистик как функций не пересекаются. Поскольку  $\varphi_0$  и  $\varphi_1$  при любых объемах наблюдений тесты уровня  $\alpha_0$  и  $\alpha_1$  соответственно, то  $P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0) = \alpha_0$  и  $P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_1) = \alpha_1$ .

$$\begin{aligned} & P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0 \cap A_1) = \\ & = \sum_{t_0, t_1} P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0 \cap A_1 \mid T_0 = t_0, T_1 = t_1) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_0 = t_0, T_1 = t_1) \end{aligned}$$

Отметим, что  $P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0 \cap A_1 \mid T_0 = t_0, T_1 = t_1) = \varphi_0(t_0)\varphi_1(t_1)$ , поскольку для того, чтобы при известных значениях статистик  $t_0$  и  $t_1$  отвергнуть гипотезы  $h_0$  и  $h_1$  в рандомизированном тесте нужно провести два испытания с вероятностью успеха  $\varphi_0(t_0)$  и  $\varphi_1(t_1)$ . Постулируется, что такие испытания независимые. Тогда:

$$\begin{aligned} P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0 \cap A_1) & = \sum_{t_0, t_1} \varphi_0(t_0)\varphi_1(t_1) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_0 = t_0, T_1 = t_1) = \\ & = \sum_{t_0, t_1} \varphi_0(t_0)\varphi_1(t_1) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_0 = t_0) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_1 = t_1) = \end{aligned}$$

$$\begin{aligned}
&= \sum_{t_0} [\varphi_0(t_0) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_0 = t_0) \sum_{t_1} \varphi_1(t_1) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_1 = t_1)] = \\
&= \sum_{t_0} [\varphi_0(t_0) P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(T_0 = t_0) \alpha_1] = \alpha_0 \alpha_1
\end{aligned}$$

По формуле полной вероятности и полного математического ожидания:

$$P_{h_0 \cap h_1}(A_0) = \sum_{\mathbf{z}} P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0) P_{h_0 \cap h_1}(\mathbf{Z} = \mathbf{z}) = \sum_{\mathbf{z}} \alpha_0 P_{h_0 \cap h_1}(\mathbf{Z} = \mathbf{z}) = \alpha_0$$

Аналогично,  $P_{h_0 \cap h_1}(A_1) = \alpha_1$ .

$$\begin{aligned}
P_{h_0 \cap h_1}(A_0 \cap A_1) &= \sum_{\mathbf{z}} P_{h_0 \cap h_1, \mathbf{Z}=\mathbf{z}}(A_0 \cap A_1) P_{h_0 \cap h_1}(\mathbf{Z} = \mathbf{z}) = \\
&= \sum_{\mathbf{z}} \alpha_0 \alpha_1 P_{h_0 \cap h_1}(\mathbf{Z} = \mathbf{z}) = \alpha_0 \alpha_1
\end{aligned}$$

Таким образом,  $P_{h_0 \cap h_1}(A_0 \cap A_1) = P_{h_0 \cap h_1}(A_0) P_{h_0 \cap h_1}(A_1) = \alpha_0 \alpha_1$ .  $\square$

В частности, положим  $h_0: X \perp\!\!\!\perp Y$  при условии  $Z = 0$ ,  $h_1: X \perp\!\!\!\perp Y$  при условии  $Z = 1$ . Тогда гипотеза проверки условной независимости будет иметь вид  $h = h_0 \cap h_1$  и тест проверки условной независимости будет иметь вид:

$$\varphi^{\text{subsamples}} = \begin{cases} 1, & \text{наступило событие } A_0 \text{ или } A_1 \\ 0, & \text{иначе} \end{cases}$$

Пусть для простоты  $\alpha_0 = \alpha_1 = \beta$ . Тогда

$$\begin{aligned}
P_{h_0 \cap h_1}(\varphi^{\text{subsamples}} = 1) &= P_{h_0 \cap h_1}(A_0 \cup A_1) = \\
&= P_{h_0 \cap h_1}(A_0) + P_{h_0 \cap h_1}(A_1) - P_{h_0 \cap h_1}(A_0 \cap A_1) = 2\beta - \beta^2
\end{aligned}$$

Нетрудно проверить, что для контроля  $P_{h_0 \cap h_1}(\varphi^{\text{subsamples}} = 1)$  достаточно положить  $\beta = 1 - \sqrt{1 - \alpha}$ .

## Список литературы

1. *Anderson T.* An Introduction to Multivariate Statistical Analysis. — Wiley-Interscience, 2003.
2. *Cramér H.* Mathematical methods of statistics. — Princeton University Press, 1946.
3. *Dai B., Ding S., Wahba G.* Multivariate Bernoulli distribution // Bernoulli. — 2013. — Т. 19, № 4. — С. 1465—1483.
4. *Lauritzen S. L.* Graphical models. — Clarendon Press, 1996.
5. *Lehmann E. L.* Testing statistical hypotheses. — Wiley, 1986.
6. *Teugels J. L.* Some representations of the multivariate Bernoulli and binomial distributions // Journal of Multivariate Analysis. — 1990. — Т. 32, № 2. — С. 256—268.