Хипергеометрично разпределение

Имаме N обекта, от които M са маркирани $(0 \le M \le N)$. Избират се n обекта и сл. вел. X е броя маркирани измежду тези N $(n \le N)$. Тогава казваме, че $X \in HG(N, M, n)$

Твърдение Нека $X \in HG(N, M, n)$. Тогава:

а) $$P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$, като $\max(0, n-(N-M)) \le k \le \min(n, M)$

б) $$\mathbb{E}[X] = n \cdot \frac{M}{N}, \quad D[X] = n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}$$

Доказателство

а) $$P(X=k) = \frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}$$

б) $X = \sum_{i=1}^{n} X_i$, където $X_i = \begin{cases} 1, & \text{ако на } i\text{-тата позиция има маркиран обект} \\ 0, & \text{ако на } i\text{-тата позиция няма маркиран обект} \end{cases}$

$$X_i \in Ber(p_i), \quad p_i = P(X_i = 1) = \frac{\binom{N-1}{M-1}}{\binom{N}{M}} = \frac{M}{N}$$

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot \frac{M}{N}$$

Съвместно разпределение на дискретни случайни величини

Деф.| (Съвм. разпр.)

Нека (X, Y) са две дискр. сл. вел. Тогава таблицата по-долу се нарича съвм. разпр. на X и Y.

$y \backslash x$	x_1	x_2	...	...	x_n	Маргинално на Y
y_1	p_{11}	p_{21}	...	...	p_{n1}	$\sum_i p_{i1}$
y_2	p_{12}	p_{22}	...	...	p_{n2}	$\sum_i p_{i2}$
...	...	...	...	...	...	...
...	...	...	...	...	...	...
y_k	p_{1k}	p_{2k}	...	...	p_{nk}	$\sum_i p_{ni}$
	$\sum_j p_{1j}$	$\sum_j p_{2j}$	...	...	$\sum_j p_{nj}$	

маргинално на X

$$\mathbb{P}(X = x_k) = \sum_j \mathbb{P}(X = x_k, Y = y_j)$$

Деф.| (Функция на разпределение на сл. вел)

Нека X и Y са произв. сл. вел. във вер. пр-во V.

Тогава ф-ята $F_{X,Y}(x,y) := \mathbb{P}(X \leq x, Y \leq y)$, $\forall (x,y) \in \mathbb{R}^2$, се нарича съвместна ф-я на разпр. на (X, Y)

$F_X(x) = F_{X,Y}(x, \infty)$ – маргинална ф-я на разпр

$F_Y(y) = F_{X,Y}(\infty, y)$

Деф.| (Независимост)

Случайните величини X и Y, дефинирани в едно вер. пр-во, са независими ($X \perp\!\!\!\perp Y$), ако $F_{X,Y}(x,y) = F_X(x) F_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$

Заб.| Ако X и Y са дискретни, горната деф. е еквивалентна на

$\mathbb{P}(X = x_k, Y = y_j) = \mathbb{P}(X = x_k)\mathbb{P}(Y = y_j)$, $\forall$ възм. ст-ти x_k и y_j

Ковариация на X и Y

Линейна зависимост $Y = a.X + b$

Деф. (Ковариация)

Нека X и Y са сл. вел. и $\mathbb{D}[X] < \infty$ и $\mathbb{D}[Y] < \infty$.

Тогава $cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ се нарича ковариация.

Твърдение $cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Доказателство $cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] =$

$$= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] =$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] =$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Следствие Ако $X \perp\!\!\!\perp Y$, то $cov(X,Y) = 0$

Доказателство Ако $X \perp\!\!\!\perp Y$, то $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \Rightarrow \underbrace{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}_{Cov(X,Y)} = 0$

Деф. (Корелация)

Нека X и Y са две сл. вел. във вер. пр-во V и $\mathbb{D}[X] < \infty$, $\mathbb{D}[Y] < \infty$,

Тогава $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{\mathbb{D}[X]}\sqrt{\mathbb{D}[Y]}}$ се нарича коеф. на корелация

Целта на корелацията е да намери степен на линейност м/у X и Y.

Тв. Нека X и Y са две сл. вел. във вер. пр-во V и $\mathbb{D}[X] < \infty$, $\mathbb{D}[Y] < \infty$.

Тогава, ако поставим $\bar{X} = \dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}}$ и $\bar{Y} = \dfrac{Y - \mathbb{E}[Y]}{\sqrt{\mathbb{D}[Y]}}$ е вярно следното:

1. $\mathbb{E}[\bar{X}] = \mathbb{E}[\bar{Y}] = 0$
2. $\mathbb{D}[\bar{X}] = \mathbb{D}[\bar{Y}] = 1$

} центриране и нормиране

3. $\rho(X,Y) = \mathbb{E}[\bar{X}.\bar{Y}]$

Доказателство

1. $\mathbb{E}[\bar{X}] = \mathbb{E}\left[\dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}}\right] = \dfrac{1}{\sqrt{\mathbb{D}[X]}} . \mathbb{E}[X - \mathbb{E}[X]] = \dfrac{\mathbb{E}[X] - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}} = 0 = \mathbb{E}[\bar{Y}]$

2. $\mathbb{D}[\bar{X}] = \mathbb{D}\left[\dfrac{X - \mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}}\right] = \dfrac{1}{\sqrt{\mathbb{D}[X]}^2} . \mathbb{D}[X - \underbrace{\mathbb{E}[X]}_{const}] = \dfrac{\mathbb{D}[X]}{\mathbb{D}[X]} = 1 = \mathbb{D}[\bar{Y}]$

3. $\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{D[X]}\sqrt{D[Y]}} = \frac{E[(X-E[X])(Y-E[Y])]}{\sqrt{D[X]}\sqrt{D[Y]}} = E\left[\underbrace{\frac{(X-E[X])}{\sqrt{D[X]}}}_{\bar{X}} \cdot \underbrace{\frac{(Y-E[Y])}{\sqrt{D[Y]}}}_{\bar{Y}}\right] = E[\bar{X}\bar{Y}]$

Теорема Нека X и Y са две сл. вел. и $D[X]<\infty$ и $D[Y]<\infty$. Тогава:

1) $|\rho(X,Y)| \leq 1$

2. $\exists a, b \in \mathbb{R}$, $Y = aX + b \Leftrightarrow |\rho(X,Y)| = 1$

$\rho(X,Y) = 0 \Rightarrow$ линейна независимост

Доказателство

1) $\rho(X,Y) = \frac{E[(X-E[X])(Y-E[Y])]}{\sqrt{D[X]}\sqrt{D[Y]}} = E[\bar{X}\bar{Y}]$

$0 \leq E[(\bar{X}+\bar{Y})^2] = \underbrace{E[\bar{X}^2]}_{=1} + \underbrace{E[\bar{Y}^2]}_{=1} + 2\underbrace{E[\bar{X}\bar{Y}]}_{=\rho(X,Y)} = 2(1+\rho(X,Y)) \geq 0 \Rightarrow \rho(X,Y) \geq -1 \quad (1)$

$0 \leq E[(\bar{X}-\bar{Y})^2] = \underbrace{E[\bar{X}^2]}_{=1} + \underbrace{E[\bar{Y}^2]}_{=1} - 2\underbrace{E[\bar{X}\bar{Y}]}_{=\rho(X,Y)} = 2(1-\rho(X,Y)) \geq 0 \Rightarrow \rho(X,Y) \leq 1 \quad (2)$

$\xRightarrow{(1) \text{ и } (2)} -1 \leq \rho(X,Y) \leq 1$

2) ($\Rightarrow$) $Y = a.X + b \quad / - E[Y]$

$Y - E[Y] = a.X + b - E[Y]$

$Y - E[Y] = aX + b - E[Y] + aE[X] - aE[X] \quad / : \sqrt{D[Y]}$

$\frac{Y-E[Y]}{\sqrt{D[Y]}} = \frac{a(X-E[X])}{\sqrt{D[Y]}} + \frac{a.E[X]+b-E[Y]}{\sqrt{D[Y]}}$

$\bar{Y} = a.\frac{(X-E[X])\sqrt{D[X]}}{\sqrt{D[Y]}\sqrt{D[X]}} + \underbrace{\frac{a.E[X]+b-E[Y]}{\sqrt{D[Y]}}}_{=W}$

$\bar{Y} = a.\bar{X}.\frac{\sqrt{D[X]}}{\sqrt{D[Y]}} + W = \underbrace{\frac{a\sqrt{D[X]}}{\sqrt{D[Y]}}}_{V}.\bar{X} + W \Leftrightarrow \bar{Y} = V.\bar{X} + W$

$0 = E[\bar{Y}] = E[V.\bar{X}+W] = V.\underbrace{E[\bar{X}]}_{=0} + W \Rightarrow W = 0 \Rightarrow \bar{Y} = V.\bar{X}$

$1 = D[\bar{Y}] = D[V.\bar{X}] = V^2.\underbrace{D[\bar{X}]}_{=1} \Rightarrow V = \pm 1$

$\rho(X,Y) = E[\bar{X}\bar{Y}] = V.\underbrace{E[\bar{X}^2]}_{=1} = V = \pm 1$

$\Leftarrow$) Нека $\rho(X,Y) = 1 = \mathbb{E}[\bar{X}\bar{Y}]$

$$0 \leq \mathbb{E}[(\bar{X}-\bar{Y})^2] = 2 - 2\mathbb{E}[\bar{X}\bar{Y}] = 0 \Rightarrow \bar{X} = \bar{Y}$$

Нека $\rho(X,Y) = -1 = \mathbb{E}[\bar{X}\bar{Y}]$

$$0 \leq \mathbb{E}[(\bar{X}+\bar{Y})^2] = 2 + 2\mathbb{E}[\bar{X}\bar{Y}] = 0 \Rightarrow \bar{X} = -\bar{Y}$$

където $\bar{X} = \bar{Y} \Leftrightarrow \dfrac{X-\mathbb{E}[X]}{\sqrt{\mathbb{D}[X]}} = \dfrac{Y-\mathbb{E}[Y]}{\sqrt{\mathbb{D}[Y]}} \Leftrightarrow Y - \mathbb{E}[Y] = X\dfrac{\sqrt{\mathbb{D}[Y]}}{\sqrt{\mathbb{D}[X]}} - \mathbb{E}[X].\dfrac{\sqrt{\mathbb{D}[Y]}}{\sqrt{\mathbb{D}[X]}}$

$$\Leftrightarrow Y = \underbrace{\frac{\sqrt{\mathbb{D}[Y]}}{\sqrt{\mathbb{D}[X]}}}_{a}.X \underbrace{- \mathbb{E}[X].\frac{\sqrt{\mathbb{D}[Y]}}{\sqrt{\mathbb{D}[X]}} + \mathbb{E}[Y]}_{b}$$

$$\Leftrightarrow Y = a.X + b$$