Новиков Лев. НА-03

T1. Вычислить точное распределение для $\hat{\tau}$ при условии, что переменные независимы, для $n=4$.

$$\hat{\tau} = \frac{K}{C_4^2}, \text{ где } K = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Q\big((X_i,Y_i),(X_j,Y_j)\big).$$

Здесь $Q\big((a,b),(c,d)\big) = \operatorname{sgn}\big((c-a)(d-b)\big)\cdot I_{\{(c-a)(d-b)\neq 0\}}$

Т.е. мы имеем пары $(X_1,Y_1),(X_2,Y_2),(X_3,Y_3),(X_4,Y_4)$.

Упорядочим их по возрастанию X:

$(X_{(1)},Y_1'),(X_{(2)},Y_2'),(X_{(3)},Y_3'),(X_{(4)},Y_4')$

Рассмотрим все возможные варианты:

y_1'	y_2'	y_3'	y_4'	K	$\hat{\tau}$
1	2	3	4	6	1
1	2	4	3	4	2/3
1	3	2	4	4	2/3
1	3	4	2	2	1/3
1	4	2	3	2	1/3
1	4	3	2	0	0
2	1	3	4	4	2/3
2	1	4	3	2	1/3
2	3	1	4	2	1/3
2	3	4	1	0	0
2	4	1	3	0	0
2	4	3	1	-2	-1/3

y_1	y_2	y_3	y_4	K	$\hat{\tau}$
3	1	2	4	2	1/3
3	1	4	2	0	0
3	2	1	4	0	0
3	2	4	1	-2	-1/3
3	4	1	2	-2	-1/3
3	4	2	1	-4	-2/3
4	1	2	3	0	0
4	1	3	2	-2	-1/3
4	2	1	3	-2	-1/3
4	2	3	1	-4	-2/3
4	3	1	2	-4	-2/3
4	3	2	1	-6	-1

Получаем, что в одном случае $\hat{\tau}=1$, в одном — -1, в трёх $\hat{\tau}=2/3$, в еще трех — $-2/3$, по 5 случаев на $\tau=1/3$ и $\tau=-1/3$, и 6 случаев $\tau=0$.

Поскольку все эти варианты равновероятны и независимы, получаем, что распр-е:

$$\hat{\tau} = \begin{cases} 1, & p=1/24 \\ 2/3, & p=1/8 \\ 1/3, & p=5/24 \\ 0, & p=1/4 \\ -1/3, & p=5/24 \\ -2/3, & p=1/8 \\ -1, & p=1/24 \end{cases}$$

T2. (X_1, Y_1), (X_2, Y_2) — i.i.d.

$$p_{(X,Y)}(x,y) = \frac{1}{2} y^2 e^{-x-y} \cdot I_{\{x>0 \,\&\, y>0\}}$$

i) Во-первых, $p_{(X,Y)}(x,y) = \frac{1}{2} y^2 e^{-y} \cdot I_{\{y>0\}} \cdot e^{-x} \cdot I_{\{x>0\}} =$

~~действительно, ведь~~ $= p_X(x) \cdot p_Y(y) \Rightarrow$

X и Y независимы

Проверяем условие нормировки: $\iint p_{(X,Y)}(x,y)\,dx\,dy$

$= \int\limits_0^\infty e^{-x} dx \cdot \int\limits_0^\infty \frac{1}{2} y^2 e^{-y} dy = \Gamma(1) \cdot \frac{1}{2} \cdot \Gamma(3) = 1 \cdot \frac{1}{2} \cdot 2 = 1.$

Аналогично с условиями нормировки для маргинальных плотностей: $\int\limits_0^\infty e^{-x} dx = \Gamma(1) = 1$,

$\int\limits_0^\infty \frac{1}{2} y^2 e^{-y} dy = \frac{1}{2} \cdot \Gamma(3) = \frac{1}{2} \cdot 2 = 1.$

Еще от функции плотности $p_{(X,Y)}(x,y) \geq 0 \;\; \forall x,y$, здесь это выполнено.

Следовательно, $p_{(X,Y)}(x,y)$ — действительно p.d.f.

ii) Как мы уже выяснили, X и Y независимы

В этом случае $\tau = 2\mathbb{P}((X_2 - X_1)(Y_2 - Y_1) > 0) - 1 =$

$= 2 \cdot (\mathbb{P}(X_2 - X_1 > 0) \cdot \mathbb{P}(Y_2 - Y_1 > 0) + \mathbb{P}(X_2 - X_1 < 0) \cdot \mathbb{P}(Y_2 - Y_1 < 0)) - 1 =$

$= 2 \cdot \left(\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}\right) - 1 = 2 \cdot \frac{1}{2} - 1 = 0$

Ответ: $\tau = 0$.

T3. $(S_1, \dots, S_n)$ – вектор рангов, распределенный равномерно над $n!$ перестановками $\{1, \dots, n\}$.

i) $\mathbb{E} S_i = \frac{1}{n} \cdot (1 + \dots + n) = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$, доказано

ii) $Var\, S_i = \mathbb{E} S_i^2 - (\mathbb{E} S_i)^2 = \frac{1}{n} \cdot (1^2 + \dots + n^2) - \frac{(n+1)^2}{4} =$

$= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(4n+2-3n-3)}{12} =$

$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$. Доказано.

iii) $cov(S_i, S_j) = \mathbb{E}(S_i S_j) - \mathbb{E}(S_i) \cdot \mathbb{E}(S_j)$

Находим $\mathbb{E}(S_i S_j)$. Все возможные пары (S_i, S_j) равновероятны $\left(\mathbb{P}(S_i = u, S_j = v) = \frac{1}{n(n-1)}\right)$.

Следовательно, $\mathbb{E}(S_i S_j) = \frac{1}{n(n-1)} \cdot \sum\limits_{u \neq v} uv$.

Находим $\sum\limits_{u \neq v} uv$: $\sum\limits_{u \neq v} uv = \sum\limits_{u=1}^{n} \sum\limits_{v=1}^{n} uv - \sum\limits_{w=1}^{n} w^2 =$

$= \sum\limits_{u=1}^{n} u \cdot \sum\limits_{v=1}^{n} v - \sum\limits_{w=1}^{n} w^2 = \left(\frac{n(n+1)}{2}\right)^2 - \frac{n(n+1)(2n+1)}{6} =$

$= \frac{n^2(n+1)^2}{4} - \frac{n(n+1)(2n+1)}{6} = \frac{3n(n+1) - 2 \cdot (2n+1)}{12} \cdot$

$\cdot n(n+1) = \frac{n(n+1)}{12} \cdot (3n^2 - n - 2) = \frac{n(n+1)(n-1)(3n+2)}{12}$.

Отсюда $\mathbb{E}(S_i S_j) = \frac{(n+1)(3n+2)}{12}$.

$\Rightarrow cov(S_i, S_j) = \frac{(n+1)(3n+2)}{12} - \left(\frac{n+1}{2}\right)^2 = \frac{3n^2+5n+2}{12} -$

$- \frac{3 \cdot (n^2+2n+1)}{12} = \frac{-n-1}{12} = -\frac{(n+1)}{12}$. Доказано!

T4. $\{X_i\}_{i=1}^m$, $\{Y_j\}_{j=1}^n$. $W=\sum_{j=1}^n S_j$

i) W минимально $\Rightarrow$ ранги Y минимальны — ранги $1,\ldots,n \Rightarrow W=\sum_{j=1}^n j = \frac{n(n+1)}{2}$.

W максимально $\Rightarrow$ ранги Y максимальны — ранги $m+1,\ldots,m+n \Rightarrow W=\sum_{j=1}^n (m+j) = \frac{n(n+1)}{2}+mn=$
$=\frac{n(2m+n+1)}{2}$.

Ответ: $W_{MIN}=\frac{n(n+1)}{2}$, $W_{MAX}=\frac{n(2m+n+1)}{2}$

ii) Для начала рассмотрим $x=0$:

$\mathbb{P}(W=\frac{n(n+1)}{2})=\mathbb{P}(S_1<\min_{1\le i\le m} R_i,\ldots,S_n<\min_{1\le i\le m} R_i)=$
$=\mathbb{P}(S_1<R_1,\ldots,S_1<R_m)\cdot\ldots\cdot\mathbb{P}(S_n<R_1,\ldots,S_n<R_m)=$
$=\prod_{i=1}^m\prod_{j=1}^n \mathbb{P}(S_j<R_i)=\left(\frac{1}{2}\right)^{mn}=2^{-mn}$.

$\mathbb{P}(W=\frac{n(2m+n+1)}{2})=\mathbb{P}(S_1>\max_{1\le i\le m} R_i,\ldots,S_n>\max_{1\le i\le m} R_i)=$
$=\prod_{i=1}^m\prod_{j=1}^n \mathbb{P}(S_j>R_i)=\left(\frac{1}{2}\right)^{mn}=2^{-mn}$.

$\Rightarrow \mathbb{P}(W=W_{min})=\mathbb{P}(W=W_{MAX})=\frac{1}{2^{mn}}$

$x=1$: $\mathbb{P}(W=W_{min}+1)=$
$=\mathbb{P}(S_j<\min_{1\le i\le m} R_i<\max_{1\le j\le n} S_j<R_i)=2^{-(n-1)}\cdot 2^{-1}\cdot$
$\cdot 2^{-(m-1)}\cdot 2^{-(m-1)(n-1)}=2^{-mn}=\frac{1}{2^{mn}}$.

Аналогично $\mathbb{P}(W=W_{MAX}-1)=\frac{1}{2^{mn}}$.

Допустим, что W отличается от W_{min} на x. Тогда есть несколько вариантов получить такую сумму рангов. Пример для $m=4$, $n=5$, $x=3$, чтобы было понятно:

YYYYXXXYX, YYYXYYXXX, YYXYYYXXX. Во всех этих случаях W – сумма рангов $Y_j = \frac{n(n+1)}{2}+3$.

Рассмотрим "симметричные" случаи с XYXXXYYYY, XXYXXYXYY, XXXYYYYXYY. Достаточно очевидно, что если для какой-то расстановки $W=W_{min}+x$, то для симметричной $W=W_{MAX}-x$. Причем $\forall m, n, x$.

Строгое док-во: Пусть есть ранги $R_1, \dots, R_m, S_1, \dots, S_n$, и $W=\sum_{j=1}^{n} S_j = W_{min}+x$.

Тогда для симметричной расстановки рангов $R_1', \dots, R_m', S_1', \dots, S_n'$ $W' = \sum_{j=1}^{n} S_j' =$

$$= \sum_{j=1}^{n}(n+m+1-S_j) = n(n+m+1) - \sum_{j=1}^{n} S_j =$$

$$= n(n+m+1) - W = n(n+m+1) - W_{min} - x =$$

$$= n(n+m+1) - \frac{n(n+1)}{2} - x = \frac{n(2m+n+1)}{2} - x =$$

$$= W_{MAX} - x.$$

(Т.к. ранги принимают значения от 1 до $m+n$, $S_j + S_j' = m+n+1$)

Итого $W = W_{MIN} + x$ и $W' = W_{MAX} - x$.

При этом $\mathbb{P}(R_i = r_i, S_j = s_j) =$

$\mathbb{P}(R_i = m+n+1-r_i, S_j = m+n+1-s_j)$.

А это значит, что для любого x

$$\mathbb{P}(W = W_{MIN} + x) = \prod_{\sum_{j=1}^{n} s_j = W_{MIN} + x} \left(\mathbb{P}(R_i = r_i, S_j = s_j)\right) =$$

$$= \prod_{\sum_{j=1}^{n} s_j = W_{MIN} + x} \left(\mathbb{P}(R_i = m+n+1-r_i, S_j = m+n+1-s_j)\right) =$$

$$= \prod_{\sum_{j=1}^{n} (m+n+1-s_j) = W_{MAX} - x} \left(\mathbb{P}(R_i = m+n+1-r_i, S_j = m+n+1-s_j)\right) =$$

$$= \prod_{\sum_{j=1}^{n} s_j' = W_{MAX} - x} \left(\mathbb{P}(R_i = r_i', S_j = s_j')\right) = \mathbb{P}(W = W_{MAX} - x)$$

(Подразумевается произведение по всем s_j, r_i, где $\sum_{j=1}^{n} s_j$ фиксирована и равна определенному значению). Доказано.