189/2012

Raport Badawczy

Research Report

RB/46/2012

Statystyki porządkowe z próby z rozkładu dyskretnego

A. Dembińska, B. Żogała

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2012

Statystyki porządkowe z próby z rozkładu dyskretnego

Anna Dembińska, Barbara Żogała

1 Wstęp

1.1 Cel pracy

Celem pracy jest opisanie rozkładów łącznych i warunkowych dla statystyk porządkowych dla prób z rozkładów dyskretnych. Sytuacja gdy zmienne losowe, na podstawie których tworzymy statystyki są niezależne o jednakowym rozkładzie jest opisana w artykule [1]. Dowody znajdujących się tam twierdzeń zostały dokładniej opisane w dalszej części pracy.

Głównym celem pracy jest opisanie przypadku, gdy zmienne losowe, na podstawie których tworzone są statystyki są niezależne, ale mają dowolne, być może różne rozkłady. Niektóre twierdzenia przedstawione w [1] daje się uogólnić na ten przypadek, jednak nie wszystkie są prawdziwe, co będzie pokazane.

Pierwszy rozdział poświęcony jest wprowadzeniu do zagadnienia poprzez podanie definicji i oznaczeń (podrozdział 1.2). W rozdziale 1.3 wyjaśnione zostało dlaczego sytuacja gdy rozkłady zmiennych są dyskretne jest trudniejsza od sytuacji zmiennych o ciągłym rozkładzie.

W drugim rozdziale rozważane są rozkłady łączne statystyk porządkowych zarówno w sytuacji, gdy zmienne mają jednakowe rozkłady, jak i gdy rozkłady są dowolne. Rozdział zaczyna się od opisania najprostszego przypadku, czyli znalezienia rozkładu pojedynczej statystyki porządkowej w sytuacji zmiennych o jednakowym rozkładzie (podrozdział 2.1). Następnie przechodzimy do rozkładu łącznego k pierwszych statystyk porządkowych (podrozdział 2.2), a na końcu przedstawione jest twierdzenie mówiące o rozkładzie łącznym dowolnego podzbioru statystyk pozycyjnych (podrozdział 2.4).

Trzeci rozdział opisuje pewien rozkład warunkowy i pokazane jest tam, że nie wszystkie własności które zachodzą dla zmiennych losowych o jednakowych rozkładach są prawdziwe w ogólniejszym przypadku zmiennych niezależnych, ale o dowolnych rozkładach.

1.2 Statystyki porządkowe - definicje i oznaczenia

W podrozdziale tym przedstawione zostaną potrzebne w pracy definicje i wykorzystywane oznaczenia. Zaczniemy od definicji statystyk porządkowych o których mowa w tej pracy.

DEFINICJA 1.1.
Niech $(X_1, X_2, \ldots, X_n)$ będzie n-wymiarowym wektorem losowym o składowych pochodzących z dowolnych rozkładów. Odpowiadający jej ***ciąg statystyk porządkowych (pozycyjnych)*** *$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ to ciąg złożony ze zmiennych losowych X_i zapisanych w niemalejącym porządku: najmniejsze spośród X_i to $X_{(1)}$, drugie co do wielkości to $X_{(2)}$, ..., największe z X_i to $X_{(n)}$.*

UWAGA 1.1.
W oczywisty sposób dla dowolnych rozkładów zachodzi nierówność

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)},$$

natomiast dla rozkładów absolutnie ciągłych powyższe nierówności są ostre, czyli

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}.$$

Oznacza to, że w przypadku gdy mamy do czynienia z rozkładami dyskretnymi, w próbkach mogą występować obserwacje związane, czyli mogą istnieć takie indeksy i, j, że $X_i = X_j$.

Zdefiniujemy teraz związane bloki obserwacji, z których będziemy korzystać w dalszej części pracy.

DEFINICJA 1.2.

Niech $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ *- ciąg liczb rzeczywistych. Mówimy, że podciąg* $t_{i_1} \leqslant t_{i_2} \leqslant \cdots \leqslant t_{i_k}$ *ma* r *związanych bloków o długości* z_j *dla każdego bloku, jeżeli:*

$$t_{i_1} = t_{i_2} = \ldots = t_{i_{z_1}} < t_{i_{z_1+1}} = \ldots = t_{i_{z_1+z_2}} < \ldots < t_{i_{z_1+\ldots+z_{r-1}+1}} = \ldots = t_{i_{z_1+\ldots+z_r}} = t_{i_k},$$

dla $r = 1, \ldots, k$ *oraz* $\sum_{j=1}^{r} z_j = k$.

UWAGA 1.2.

Najczęściej wykorzystywanym podciągiem będzie po prostu k pierwszych wartości ciągu $t_1 \leqslant \ldots \leqslant t_n$, czyli $t_1 \leqslant \ldots \leqslant t_k$. Będziemy go czasami oznaczać $\mathbb{t}_k = (t_1, \ldots, t_k)$.

Poniżej zaprezentujemy przykład związanego bloku obserwacji.

PRZYKŁAD 1.1.

Załóżmy, że wylosowano 30 obserwacji:

$$3,\ 0,\ 6,\ 2,\ 0,\ 8,\ 3,\ 9,\ 3,\ 6,\ 3,\ 2,\ 4,\ 6,\ 6,\ 1,\ 3,\ 2,\ 0,\ 0,\ 2,\ 8,\ 9,\ 5,\ 5,\ 5,\ 4,\ 8,\ 9,\ 9.$$

Układając je w kolejności rosnącej i odcinając pierwsze 20, otrzymamy:

$$0,\ 0,\ 0,\ 0,\ 1,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 3,\ 3,\ 3,\ 4,\ 4,\ 5,\ 5,\ 5,\ 6\,|\,6,\ 6,\ 6,\ 8,\ 8,\ 8,\ 9,\ 9,\ 9,\ 9.$$

Zatem w tym wypadku:

$$\mathbb{t}_{20} = (t_1, \ldots, t_{20}) = (0,0,0,0,1,2,2,2,2,3,3,3,3,3,4,4,5,5,5,6).$$

Otrzymaliśmy w ten sposób 7 związanych bloków o długościach:

$$\underbrace{0,0,0,0}_{z_1=4},\underbrace{1}_{z_2=1},\underbrace{2,2,2,2}_{z_3=4},\underbrace{3,3,3,3,3}_{z_4=5},\underbrace{4,4}_{z_5=2},\underbrace{5,5,5}_{z_6=3},\underbrace{6}_{z_7=1}.$$

Kolejną definicją, wykorzystywaną w dalszej części pracy do czytelnego zapisu wyprowadzonych wzorów będzie definicja permanentu.

DEFINICJA 1.3.

Niech $A = (a_{i,j})_{i,j=1,\ldots,n}$ *będzie kwadratową macierzą rzędu* n. *Wtedy permanent macierzy* A *jest zdefiniowany jako:*

$$PerA = \sum_{\substack{(j_1,\ldots,j_n)- \\ permutacje \\ zb.\ \{1,\ldots,n\}}} \prod_{i=1}^{n} a_{i,j_i}$$

UWAGA 1.3.

Definicja permanentu jest bardzo podobna do definicji wyznacznika macierzy. Różnica polega na tym, że w permanencie składniki sumy dla każdej permutacji są brane ze znakiem „+", podczas gdy w definicji wyznacznika są brane ze znakiem permutacji.

PRZYKŁAD 1.2.

Jako przykład znajdziemy wyznacznik i permanent poniższej macierzy

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}.$$

W tym celu szukamy wszystkich permutacji zbioru $\{1,2,3\}$: $(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)$. Permanent i wyznacznik wynoszą odpowiednio

$$\begin{aligned}\mathrm{Per}A &= a_{11}a_{22}a_{33} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} + a_{13}a_{22}a_{31} = \\ &= 5\cdot 9 + 8\cdot 6 + 4\cdot 2\cdot 9 + 4\cdot 8\cdot 3 + 7\cdot 2\cdot 6 + 7\cdot 5\cdot 3 = 450, \\ \mathrm{Det}A &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} = \\ &= 5\cdot 9 - 8\cdot 6 - 4\cdot 2\cdot 9 + 4\cdot 8\cdot 3 + 7\cdot 2\cdot 6 - 7\cdot 5\cdot 3 = 0.\end{aligned}$$

W pracy będziemy wykorzystywali przedstawione poniżej oznaczenia.

OZNACZENIE 1.1.
Niech $X_1,\ldots,X_n$ będą zmiennymi losowymi o dyskretnych rozkładach z dystrybuantami $F_1,\ldots,F_n$ i funkcjami prawdopodobieństwa $p_1,\ldots,p_n$ odpowiednio. Taką sytuację będziemy oznaczać $X_i \sim F_i, p_i$ dla $i=1,\ldots,n$. Jeśli dodatkowo rozkłady te będą takie same, czyli $F_i = F, p_i = p$ dla $i=1,\ldots,n$, to sytuację taką będziemy oznaczać $X_1,\ldots,X_n \sim F,p$ lub $X_i \sim F,p$ dla $i=1,\ldots,n$.

OZNACZENIE 1.2.
Niech $X_1,\ldots,X_n$ będą niezależnymi zmiennymi losowymi o takich samych dyskretnych rozkładach z dystrybuantami F i funkcjami prawdopodobieństwa p. Taką sytuację będziemy oznaczać $X_1,\ldots,X_n \overset{\text{iid}}{\sim} F,p$ lub $X_i \overset{\text{iid}}{\sim} F,p$ dla $i=1,\ldots,n$.

OZNACZENIE 1.3.
Niech F będzie dystrybuantą rozkładu dyskretnego, a $x_0 \in \mathbb{R}$. Wtedy poprzez $F(x_0^-)$ będziemy rozumieli:

$$F(x_0^-) = \lim_{x\to x_0^-} F(x).$$

1.3 Dlaczego przypadek zmiennych dyskretnych jest trudniejszy od przypadku ciągłego

Rozkłady statystyk porządkowych dla przypadku zmiennych o rozkładach absolutnie ciągłych są znane i opisane na przykład w pracy [2]. Nie można jednak tych rozważań łatwo przenieść na przypadek, gdy zmienne mają rozkłady dyskretne. Wynika to z faktu, o którym wspomniano przy definicji statystyk pozycyjnych: w próbach pochodzących z rozkładów dyskretnych mogą występować obserwacje związane.
Sytuacja zmiennych dyskretnych jest rozważana w pracy [1], jednak praca ta dotyczy jedynie przypadku, gdy wszystkie zmienne mają taki sam rozkład. W poniższej pracy poza dokładnym opisaniem przypadku z [1], zostanie rozpatrzona sytuacja zmiennych niezależnych o dowolnych rozkładach.

2 Rozkłady łączne statystyk porządkowych

Poniższy rozdział poświęcony jest rozkładom łącznym statystyk porządkowych. Rozważania rozpoczniemy od najprostszego przypadku, czyli znalezienia rozkładu pojedynczej statystyki porządkowej w sytuacji, gdy obserwacje na podstawie których tworzone są statystyki są niezależne i pochodzą z takiego samego rozkładu dyskretnego. Następnie zajmiemy się już różnymi rozkładami łącznymi w sytuacjach gdy zmienne są niezależne i mają takie same lub dowolne rozkłady.

2.1 Rozkład pojedynczej statystyki $\mathbb{P}(X_{(k)} = x)$

Na początku znajdziemy rozkład prawdopodobieństwa pojedynczej statystyki porządkowej, czyli $\mathbb{P}(X_{(k)} = x)$, w najprostszej sytuacji: $X_1,\ldots,X_n \overset{\text{iid}}{\sim} F,p$. Można to zrobić na kilka sposobów.
Pierwszym z nich jest znalezienie bezpośrednio funkcji prawdopodobieństwa.

Funkcja prawdopodobieństwa: $\mathbb{P}(X_{(k)} = x)$

Jak zauważyliśmy wcześniej, w odróżnieniu od przypadku ciągłego nie można stwierdzić, że z prawdopodobieństwem 1 zachodzi $X_{(k-1)} < X_{(k)}$ i $X_{(k+1)} > X_{(k)}$. Trzeba rozważyć również sytuacje, w których mogą występować obserwacje związane, zatem z prawdopodobieństwem większym od 0 zarówno może zachodzić $X_{(k_1)} = X_{(k)}$ dla $k_1 < k$, jak i $X_{(k_2)} = X_{(k)}$ dla $k_2 > k$. Możliwe do zaistnienia sytuacje obrazuje rysunek:

$$\underbrace{X_{(1)} \quad X_{(k-s-1)}}_{<x} \qquad \underbrace{X_{(k-s)} \quad X_{(k)} \qquad X_{(k+u)}}_{=x} \qquad \underbrace{X_{(k+u+1)} \quad X_{(n)}}_{>x}$$

$$\text{dla} \quad s = 0, \ldots, k-1 \qquad \text{oraz} \quad u = 0, \ldots, n-k$$

Zdefiniujmy zatem zdarzenia odpowiadające powyższej sytuacji:

$$A_{su} : \begin{cases} \text{dokładnie } s+u+1 \text{ obserwacji jest } = x, \\ \text{dokładnie } k-s-1 \text{ obserwacji jest } < x, \\ \text{dokładnie } n-k-u \text{ obserwacji jest } > x. \end{cases}$$

Zdarzenia te są parami rozłączne, zatem:

$$\mathbb{P}(X_{(k)} = x) = \mathbb{P}(\bigcup_{s=0}^{k-1} \bigcup_{u=0}^{n-k} A_{su}) = \sum_{s=0}^{k-1} \sum_{u=0}^{n-k} \mathbb{P}(A_{su}).$$

Żeby znaleźć $\mathbb{P}(A_{su})$, zdefiniujemy kolejne zdarzenia:

$$\begin{cases} B_{1(su)} : & \text{dokładnie } s+u+1 \text{ obserwacji jest } = x, \\ B_{2(su)} : & \text{dokładnie } k-s-1 \text{ obserwacji jest } < x, \\ B_{3(su)} : & \text{dokładnie } n-k-u \text{ obserwacji jest } > x, \end{cases}$$

wtedy:

$$\begin{aligned} \mathbb{P}(A_{su}) &= \mathbb{P}(B_{1(su)}, B_{2(su)}, B_{3(su)}) = \\ &= \mathbb{P}(B_{1(su)})\mathbb{P}(B_{2(su)}|B_{1(su)})\mathbb{P}(B_{3(su)}|B_{1(su)}, B_{2(su)}) = \\ &= \binom{n}{s+u+1}[p(x)]^{s+u+1}\binom{n-s-u-1}{k-s-1}[F(x^-)]^{k-s-1}[1-F(x)]^{n-k-u} = \\ &= \frac{n!}{(s+u+1)!\cancel{(n-s-u-1)!}} \frac{\cancel{(n-s-u-1)!}}{(k-s-1)!(n-u-k)!}[p(x)]^{s+u+1}[F(x^-)]^{k-s-1}[1-F(x)]^{n-k-u} = \\ &= \frac{n!}{(s+u+1)!(k-s-1)!(n-u-k)!}[p(x)]^{s+u+1}[F(x^-)]^{k-s-1}[1-F(x)]^{n-k-u}. \end{aligned}$$

Zatem ostatecznie:

$$\mathbb{P}(X_{(k)} = x) = \sum_{s=0}^{k-1} \sum_{u=0}^{n-k} \frac{n!}{(s+u+1)!(k-s-1)!(n-u-k)!}[p(x)]^{s+u+1}[F(x^-)]^{k-s-1}[1-F(x)]^{n-k-u}. \tag{2.1}$$

Dystrybuanta rozkładu: $\mathbb{P}(X_{(k)} \leqslant x)$

Drugim sposobem jest znalezienie $\mathbb{P}(X_{(k)} \leqslant x)$, czyli dystrybuanty $F_{X_{(k)}}(x)$, którą oznaczymy $F_{(k)}(x)$.

$$\mathbb{P}(X_{(k)} \leqslant x) = \mathbb{P}(\text{co najmniej } k \text{ obserwacji jest } \leqslant x) = (*).$$

Zdefiniujmy zatem zdarzenia dla $s = 0, \ldots, n-k$:

$$A_s : \begin{cases} \text{dokładnie } k+s \text{ obserwacji jest } \leqslant x & \text{oznaczmy to jako } B_{1(s)}, \\ \text{dokładnie } n-k-s \text{ obserwacji jest } > x & \text{oznaczmy to jako } B_{2(s)}. \end{cases}$$

Zdarzenia A_s są parami rozłączne, zatem:

$$\begin{aligned}(*) &= \mathbb{P}(\bigcup_{s=0}^{n-k} A_s) = \sum_{s=0}^{n-k} \mathbb{P}(A_s) = \\ &= \sum_{s=0}^{n-k} \mathbb{P}(B_{1(s)})\mathbb{P}(B_{2(s)}|B_{1(s)}) = \\ &= \sum_{s=0}^{n-k} \binom{n}{k+s} [F(x)]^{k+s}[1-F(x)]^{n-k-s}.\end{aligned}$$

Jeżeli chcemy jednak znaleźć $\mathbb{P}(X_{(k)} = x)$, to wystarczy skorzystać z faktu, że:

$$\begin{aligned}&\mathbb{P}(X_{(k)} = x) = F_{(k)}(x) - F_{(k)}(x^-) = \\ &= \sum_{s=0}^{n-k} \binom{n}{k+s} \Big([F(x)]^{k+s}[1-F(x)]^{n-k-s} - [F(x^-)]^{k+s}[1-F(x^-)]^{n-k-s} \Big).\end{aligned}$$

W ten sposób znaleźliśmy rozkład pojedynczej, k-tej statystyki porządkowej. W dalszej części tego rozdziału zajmiemy się rozkładami łącznymi statystyk porządkowych.

2.2 Rozkład łączny k pierwszych statystyk porządkowych $\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k)$

2.2.1 Przypadek zmiennych niezależnych o jednakowym rozkładzie

Pierwszym i najprostszym rozkładem łącznym jaki znajdziemy będzie rozkład k pierwszych statystyk porządkowych. Ponieważ tak jak w powyższym podrozdziale możliwe jest występowanie obserwacji związanych, skorzystamy z notacji wprowadzonej w podrozdziale 1.2 i niech $t_1, \ldots, t_n$ będą wartościami realizacji próby. Zakładamy, że

$$t_1 = t_2 = \ldots = t_{z_1} < t_{z_1+1} = \ldots = t_{z_1+z_2} < \ldots < t_{z_1+\ldots+z_{r-1}+1} = \ldots = t_{z_1+\ldots+z_r} = t_k.$$

Oznacza to, że $t_1, \ldots, t_k$ ma r związanych bloków, zatem wystąpiło dokładnie r różnych obserwacji i każda z nich wystąpiła odpowiednio $z_1, \ldots, z_r$ razy. Przy tych założeniach postać rozkładu opisuje poniższe twierdzenie.

TWIERDZENIE 2.1.
Niech $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$.
Wtedy:

$$\begin{aligned}&\mathbb{P}(\mathbb{X}_{(k)} = \mathbb{t}_k) = \mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \\ &= \sum_{s=0}^{n-k} \frac{n!}{(n-k-s)!(z_r+s)!\prod_{j=1}^{r-1} z_j!} [\prod_{i=1}^{k} p(t_i)][p(t_k)]^s[1-F(t_k)]^{n-k-s}.\end{aligned} \tag{2.2}$$

Dowód:

Szukane prawdopodobieństwo można zapisać w następujący sposób:

$$\begin{aligned}
\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \mathbb{P}(&z_1 \text{ obserwacji jest równych } t_{z_1}, \\
&z_2 \text{ obserwacji jest równych } t_{z_1+z_2}, \\
&\ldots, \\
&z_{r-1} \text{ obserwacji jest równych } t_{z_1+\cdots+z_{r-1}}, \\
&\text{co najmniej } z_r \text{ obserwacji jest równych } t_k, \\
&\text{pozostałe są większe od } t_k).
\end{aligned}$$

Interesującą nas sytuację obrazuje poniższy rysunek:

$$\underbrace{X_{(1)} \cdots X_{(z_1)}}_{=t_{z_1}} \quad \cdots \quad \underbrace{X_{(z_1+\cdots+z_{r-1}+1)} \cdots X_{(k)}}_{=t_k} \; \underbrace{X_{(k+1)} \cdots X_{(k+s)}}_{=t_k} \; \underbrace{X_{(k+s+1)} \cdots X_{(n)}}_{>t_k}$$

Zdefiniujmy zatem odpowiednie zdarzenia dla możliwych wartości s, czyli dla $s = 1, \ldots, n-k$:

$$A_s : \begin{cases} z_1 & \text{obserwacji jest równych } t_1 = \ldots = t_{z_1}, \\ z_2 & \text{obserwacji jest równych } t_{z_1+1} = \ldots = t_{z_1+z_2}, \\ \vdots & \\ z_r + s & \text{obserwacji jest równych } t_{z_1+\ldots+z_r} = t_k, \\ n-k-s & \text{obserwacji jest większych od } t_k. \end{cases}$$

Wtedy, korzystając z tego, że zdarzenia A_s są parami rozłączne otrzymujemy:

$$\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \mathbb{P}(\bigcup_{s=0}^{n-k} A_s) = \sum_{s=0}^{n-k} \mathbb{P}(A_s) = (*).$$

Aby znaleźć prawdopodobieństwo $\mathbb{P}(A_s)$ zdefiniujemy kolejne zdarzenia:

$$\begin{cases} B_1 : & z_1 & \text{obserwacji jest równych } t_1 = \cdots = t_{z_1}, \\ B_2 : & z_2 & \text{obserwacji jest równych } t_{z_1+1} = \cdots = t_{z_2}, \\ \vdots & & \\ B_{r(s)} : & z_r + s & \text{obserwacji jest równych } t_{z_1+\cdots+z_r} = t_k, \\ C_s : & n-k-s & \text{obserwacji jest większych od } t_k. \end{cases}$$

Zatem $A_s = \bigcap_{l=1}^{r-1} B_l \cap B_{r(s)} \cap C_{(s)}$, więc możemy znaleźć szukane prawdopodobieństwo:

$$
\begin{aligned}
(*) &= \sum_{s=0}^{n-k} \mathbb{P}(B_1, \ldots, B_{r(s)}, C_s) = \sum_{s=0}^{n-k} \mathbb{P}(B_1)\mathbb{P}(B_2|B_1) \cdot \ldots \cdot \mathbb{P}(B_{r(s)}|B_1, \ldots, B_{r-1})\mathbb{P}(C_s|B_1, \ldots, B_{r(s)}, C_s) = \\
&= \sum_{s=0}^{n-k} \binom{n}{z_1} \underbrace{[p(t_{z_1})]^{z_1}}_{\prod_{j=1}^{z_1} p(t_j)} \binom{n-z_1}{z_2} \underbrace{[p(t_{z_1+z_2})]^{z_2}}_{\prod_{j=z_1+1}^{z_1+z_2} p(t_j)} + \ldots + \binom{n-z_1-\ldots-z_{r-1}}{z_r+s} \underbrace{[p(t_{z_1+\ldots+z_r})]^{z_r+s}}_{\prod_{j=z_1+\ldots+z_{r-1}+1}^{k+s} p(t_j)} [1-F(t_k)]^{n-k-s} = \\
&= \sum_{s=0}^{n-k} \frac{n!}{z_1!\cancel{(n-z_1)!}} \frac{\cancel{(n-z_1)!}}{\cancel{(n-z_1-z_2)!}z_2!} \cdot \ldots \cdot \frac{\cancel{(n-z_1-\ldots-z_{r-1})!}}{\underbrace{(n-z_1-\ldots-z_r-s)!}_{(n-k-s)!}(z_r+s)!} \prod_{j=1}^{k} p(t_j)[p(t_k)]^s[1-F(t_k)]^{n-k-s} = \\
&= \sum_{s=0}^{n-k} \frac{n!}{\prod_{j=1}^{r-1} z_j!(n-k-s)!(z_r+s)!} \prod_{j=1}^{k} p(t_j)[p(t_k)]^s[1-F(t_k)]^{n-k-s}.
\end{aligned}
$$

□

2.2.2 Przypadek zmiennych niezależnych o dowolnych rozkładach

Powyższe twierdzenie 2.1 dotyczy sytuacji, w której wszystkie obserwacje X_i są niezależne i mają taki sam rozkład. Możemy się jednak zastanowić jak zmieni się przedstawiony wzór, jeżeli przyjmiemy, że jedna spośród zmiennych X_i ma inny rozkład od pozostałych. Sytuację tę można opisać przy pomocy poniższego twierdzenia.

TWIERDZENIE 2.2.
Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi oraz: $X_1 \sim G, g$ i $X_2, \ldots, X_n \overset{\text{iid}}{\sim} F, p$.
Wtedy:

$$
\begin{aligned}
&\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \\
&= \frac{(n-1)!}{\prod_{j=1}^{r-1} z_j!} \prod_{j=1}^{k} p(t_j) \sum_{s=0}^{n-k} \left[\frac{[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(z_r+s)!(n-k-s)!} \left(\sum_{m=1}^{r-1} \frac{z_m g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} + \frac{(z_r+s)g(t_k)}{p(t_k)} + \right.\right. \\
&\qquad \left.\left. + \frac{(n-k-s)[1-G(t_k)]}{[1-F(t_k)]} \right) \right]. \qquad (2.3)
\end{aligned}
$$

Dowód:
Na początku definiujemy zdarzenia A_s oraz $B_1, \ldots, B_{r(s)}, C_{(s)}$ tak samo jak w dowodzie twierdzenia 2.1:

$$
A_s : \begin{cases}
z_1 & \text{obserwacji jest równych } t_1 = \ldots = t_{z_1} \quad (\text{ozn. } B_1), \\
z_2 & \text{obserwacji jest równych } t_{z_1+1} = \ldots = t_{z_1+z_2} \quad (\text{ozn. } B_2), \\
\vdots & \\
z_r + s & \text{obserwacji jest równych } t_{z_1+\ldots+z_r} = t_k \quad (\text{ozn. } B_{r(s)}), \\
n-k-s & \text{obserwacji jest większych od } t_k \quad (\text{ozn. } C_s),
\end{cases}
$$

i otrzymujemy:

$$
\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \sum_{s=0}^{n-k} \mathbb{P}(A_s) = \sum_{s=0}^{n-k} \mathbb{P}(B_1, \ldots, B_{r(s)}, C_{(s)}) = (*).
$$

Następnie zdefiniujmy zdarzenia D_m dla $m = 1, \ldots, r$ oraz D i $D*$, odpowiadające wszystkim możliwym wartościom zmiennej X_1.

$$\begin{cases} D_m: & X_1 = t_{z_1+\ldots+z_m} \text{ gdzie } m = 1, \ldots, r, \\ D: & X_1 \neq t_{z_1+\ldots+z_m} \ \forall m = 1, \ldots, r \wedge X_1 < t_k, \\ D^*: & X_1 > t_k. \end{cases}$$

Wtedy $D_1, \ldots, D_r$, D, D^* są parami rozłączne oraz $\mathbb{P}(\bigcup_{m=1}^{r} D_m \cup D \cup D^*) = 1$. Oznacza to, że możemy skorzystać ze wzoru na prawdopodobieństwo całkowite, warunkując po zdarzeniach D_m, D, D^*. W ten sposób rozważymy osobno każdą możliwą wartość X_1.

$$(*) = \sum_{s=0}^{n-k} \left[\sum_{m=1}^{r} \underbrace{\mathbb{P}(B_1, \ldots, C_{(s)}|D_m)\mathbb{P}(D_m)}_{(1*m)} + \underbrace{\mathbb{P}(B_1, \ldots, C_{(s)}|D)\mathbb{P}(D)}_{(2^*)} + \underbrace{\mathbb{P}(B_1, \ldots, C_{(s)}|D^*)\mathbb{P}(D^*)}_{(3^*)} \right] = (**).$$

Rozpiszemy teraz każdy ze składników sumy osobno.
Pierwsza grupa składników sumy dla $m = 1, \ldots, r-1$ wynosi:

$$\begin{aligned}(1*m) =& \binom{n-1}{z_1}[p(t_{z_1})]^{z_1} \cdot \ldots \cdot \binom{n - z_1 - \ldots - z_{m-1} - 1}{z_m - 1}[p(t_{z_1+\ldots+z_m})]^{z_m-1} g(t_{z_1+\ldots+z_m}) \times \\ &\times \binom{n - z_1 - \ldots - z_m}{z_{m+1}}[p(t_{z_1+\ldots+z_{m+1}})]^{z_{m+1}} \cdot \ldots \cdot \binom{n - z_1 - \ldots - z_{r-1}}{z_r + s}[p(t_k)]^{z_r+s} \times \\ &\times \binom{n-k-s}{n-k-s}[1 - F(t_k)]^{n-k-s} = \\ =& \frac{(n-1)! z_m}{\prod_{j=1}^{r-1} z_j!(z_r+s)!(n-k-s)!} \prod_{j=1}^{k} p(t_j)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} \frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})},\end{aligned}$$

a dla $m = r$:

$$\begin{aligned}(1*r) =& \binom{n-1}{z_1}[p(t_{z_1})]^{z_1} \cdot \ldots \cdot \binom{n - z_1 - \ldots - z_{r-1} - 1}{z_r + s - 1}[p(t_k)]^{z_r+s-1} g(t_k) \binom{n-k-s}{n-k-s}[1 - F(t_k)]^{n-k-s} = \\ =& \frac{(n-1)!(z_r+s)}{\prod_{j=1}^{r-1} z_j!(z_r+s)!(n-k-s)!} \prod_{j=1}^{k} p(t_j)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} \frac{g(t_k)}{p(t_k)}.\end{aligned}$$

Ponieważ $\mathbb{P}(B_1, \ldots, C_{(s)}|D) = 0$, to $(2^*) = 0$, a trzeci składnik sumy wynosi:

$$\begin{aligned}(3^*) =& \binom{n-1}{z_1}[p(t_{z_1})]^{z_1} \cdot \ldots \cdot \binom{n - z_1 - \ldots - z_{r-1} - 1}{z_r + s}[p(t_k)]^{z_r+s} \binom{n-k-s-1}{n-k-s-1}[1 - F(t_k)]^{n-k-s-1}[1 - G(t_k)] = \\ =& \frac{(n-1)!(n-k-s)}{\prod_{j=1}^{r-1} z_j!(z_r+s)!(n-k-s)!} \prod_{j=1}^{k} p(t_j)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} \frac{1 - G(t_k)}{1 - F(t_k)}.\end{aligned}$$

Zatem:

$$(**) = \sum_{s=0}^{n-k}\left[\sum_{m=1}^{r-1}\frac{(n-1)!z_m}{\prod_{j=1}^{r-1} z_j!(z_r+s)!(n-k-s)!}\prod_{j=1}^{k}p(t_j)[p(t_k)]^s[1-F(t_k)]^{n-k-s}\frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})}+\right.$$

$$+\frac{(n-1)!(z_r+s)}{\prod_{j=1}^{r-1} z_j!(z_r+s)!(n-k-s)!}\prod_{j=1}^{k}p(t_j)[p(t_k)]^s[1-F(t_k)]^{n-k-s}\frac{g(t_k)}{p(t_k)}+$$

$$\left.+\frac{(n-1)!(n-k-s)}{\prod_{j=1}^{r-1} z_j!(z_r+s)!(n-k-s)!}\prod_{j=1}^{k}p(t_j)[p(t_k)]^s[1-F(t_k)]^{n-k-s}\frac{1-G(t_k)}{1-F(t_k)}\right]=$$

$$=\frac{(n-1)!}{\prod_{j=1}^{r-1} z_j!}\prod_{j=1}^{k}p(t_j)\sum_{s=0}^{n-k}\left[\frac{[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(z_r+s)!(n-k-s)!}\left(\sum_{m=1}^{r-1}\frac{z_m g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})}+\frac{(z_r+s)g(t_k)}{p(t_k)}+\right.\right.$$

$$\left.\left.+\frac{(n-k-s)[1-G(t_k)]}{[1-F(t_k)]}\right)\right].$$

□

Kolejnym uogólnieniem będzie znalezienie powyższego prawdopodobieństwa w sytuacji, gdy zmienne są niezależne, jednak każda ze zmiennych X_i ma inny rozkład F_i.

TWIERDZENIE 2.3.

Niech $X_i \sim F_i, p_i$ dla $i = 1, \ldots, n$ oraz zmienne X_i są niezależne.
Wtedy:

$$\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) =$$

$$=\sum_{s=0}^{n-k}\sum_{\substack{(j_1,\ldots,j_n)\text{ - }\\ \text{permutacje}\\ \text{zb. }\{1,\ldots,n\}}}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+s)!(n-k-s)!}\prod_{i=1}^{k}p_{j_i}(t_i)\prod_{i=k+1}^{k+s}p_{j_i}(t_k)\prod_{i=k+s+1}^{n}[1-F_{j_i}(t_i)]=$$

$$=\sum_{s=0}^{n-k}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+s)!(n-k-s)!}PerM_s, \tag{2.4}$$

gdzie dla $s = 1, \ldots, n-k$ macierze M_s mają postać:

$$M_s = \begin{bmatrix} p_1(t_1) & p_2(t_1) & \ldots & p_n(t_1) \\ p_1(t_2) & p_2(t_2) & \ldots & p_n(t_2) \\ \vdots & \vdots & & \vdots \\ p_1(t_k) & p_2(t_k) & \ldots & p_n(t_k) \\ \vdots & \vdots & & \vdots \\ p_1(t_k) & p_2(t_k) & \ldots & p_n(t_k) \\ 1-F_1(t_k) & 1-F_2(t_k) & \ldots & 1-F_n(t_k) \\ \vdots & \vdots & & \vdots \\ 1-F_1(t_k) & 1-F_2(t_k) & \ldots & 1-F_n(t_k) \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\\a\\a\\a\end{matrix}}\right\} k+s \textit{ wierszy} \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n-k-s \textit{ wierszy} \end{array}$$

Dowód:

Ponownie szukamy prawdopodobieństwa

$$\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \sum_{s=0}^{n-k}\mathbb{P}(z_1 \text{ obserwacji } = t_{z_1}, \ldots, z_r{+}s \text{ obserwacji } = t_k, n{-}k{-}s \text{ obserwacji } > t_k) = (*)$$

Jednak tym razem każda ze zmiennych X_i ma inny rozkład, zatem nie można zliczać ich tak jak w poprzednim przypadku.

Zdefiniujmy zdarzenia:

$$P_{(j_1,\dots,j_n)}:\begin{cases} X_{j_1}=\ldots=X_{j_{z_1}}=t_{z_1},\\ \vdots \\ X_{j_{z_1+\ldots+z_{r-1}+1}}=\ldots=X_{j_{k+s}}=t_k,\\ X_{j_{k+s+1}}>t_k,\ldots,X_{j_n}>t_k,\end{cases} \qquad \text{gdzie } (j_1,\dots,j_n) \text{ - dowolna permutacja } \{1,\dots,n\}.$$

Wtedy:

$$(*)=\sum_{s=0}^{n-k}\mathbb{P}(\bigcup_{(j_1,\dots,j_n)}P_{(j_1,\dots,j_n)})=(**).$$

Suma $\mathbb{P}(\bigcup_{(j_1,\dots,j_n)}P_{(j_1,\dots,j_n)})$ nie składa się jednak z rozłącznych zdarzeń, ponieważ istnieją różne permutacje: $(j_1,\dots,j_n)\neq(k_1,\dots,k_n)$ tż: $P_{(j_1,\dots,j_n)}=P_{(k_1,\dots,k_n)}$.
np. $(1,2,\dots,z_1,\dots,n)\neq(2,1,\dots,z_1,\dots,n)$, ale odpowiadające im zdarzenia są równe.
Trzeba się zatem zastanowić ile razy każde zdarzenie występuje w powyższej sumie. Aby je zliczyć spojrzymy na postać dowolnego z nich:

$$\underbrace{X_{j_1},\dots,X_{j_{z_1}}}_{\substack{\text{permutacja takiej grupy}\\ \text{nie zmieni zdarzenia,}\\ \text{gdyż każdy z X-ów z tej}\\ \text{grupy jest } =t_{z_1}.\\ \text{grupę można spermutować}\\ \text{na } z_1! \text{ sposobów}}},\ \dots,\ \underbrace{X_{j_{z_1+\dots+z_{r-1}+1}},\dots,X_{j_{z_1+\dots+z_r+s}}}_{(z_r+s)!\ \text{permutacji}},\underbrace{X_{j_{k+s+1}},\dots,X_{j_n}}_{(n-k-s)!\text{permutacji}}$$

Rozważając analogicznie pozostałe grupy obserwacji otrzymamy, że każde zdarzenie powtarza się w sumie $\bigcup_{(j_1,\dots,j_n)}P_{(j_1,\dots,j_n)}$ dokładnie $z_1!\cdot\ldots\cdot z_{r-1}!(z_r+s)!(n-k-s)!$ razy. Zatem prawdziwa będzie równość:

$$\mathbb{P}(\bigcup_{(j_1,\dots,j_n)}P_{(j_1,\dots,j_n)})=\sum_{(j_1,\dots,j_n)}\frac{1}{z_1!\cdot\ldots\cdot z_{r-1}!(z_r+s)!(n-k-s)}\mathbb{P}(P_{(j_1,\dots,j_n)}).$$

Pozostaje znalezienie $\mathbb{P}(P_{(j_1,\dots,j_n)})$, które łatwo wyznaczymy korzystając z niezależności obserwacji $X_1,\dots,X_n$.

$$\begin{aligned}\mathbb{P}(P_{(j_1,\dots,j_n)})=&\\ &=\mathbb{P}(X_{j_1}=\ldots=X_{j_{z_1}}=t_{z_1},\dots,X_{j_{z_1+\ldots+z_{r-1}+1}}=\ldots=X_{j_{k+s}}=t_k,X_{j_{k+s+1}}>t_k,\dots,X_{j_n}>t_k)=\\ &=\prod_{i=1}^{k}p_{j_i}(t_i)\prod_{i=k+1}^{k+s}p_{j_i}(t_k)\prod_{i=k+s+1}^{n}[1-F_{j_i}(t_k)].\end{aligned}$$

Ostatecznie otrzymujemy:

$$\begin{aligned}(**)&=\sum_{s=0}^{n-k}\mathbb{P}(\bigcup_{(j_1,\dots,j_n)}P_{(j_1,\dots,j_n)})=\\ &=\sum_{s=0}^{n-k}\sum_{(j_1,\dots,j_n)}\frac{1}{z_1!\cdot\ldots\cdot z_{r-1}!(z_r+s)!(n-k-s)!}\prod_{i=1}^{k}p_{j_i}(t_i)\prod_{i=k+1}^{k+s}p_{j_i}(t_k)\prod_{i=k+s+1}^{n}[1-F_{j_i}(t_k)]=(***).\end{aligned}$$

Jeżeli dodatkowo dla $s = 1, \ldots, n-k$ zdefiniujemy macierze

$$M_s = \left[\begin{array}{cccc} p_1(t_1) & p_2(t_1) & \ldots & p_n(t_1) \\ p_1(t_2) & p_2(t_2) & \ldots & p_n(t_2) \\ \vdots & \vdots & & \vdots \\ p_1(t_k) & p_2(t_k) & \ldots & p_n(t_k) \\ \vdots & \vdots & & \vdots \\ p_1(t_k) & p_2(t_k) & \ldots & p_n(t_k) \\ 1-F_1(t_k) & 1-F_2(t_k) & \ldots & 1-F_n(t_k) \\ \vdots & \vdots & & \vdots \\ 1-F_1(t_k) & 1-F_2(t_k) & \ldots & 1-F_n(t_k) \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} a\\a\\a\\a\\a\\a \end{array}}\right\} k+s \text{ wierszy} \\ \left.\vphantom{\begin{array}{c} a\\a\\a \end{array}}\right\} n-k-s \text{ wierszy} \end{array}$$

to przy pomocy definicji permanentu (definicja 1.3) możemy uprościć zapis:

$$(***) = \sum_{s=0}^{n-k} \frac{1}{\prod\limits_{i=1}^{r-1} z_i!(z_r+s)!(n-k-s)!} \mathrm{Per} M_s.$$

□

UWAGA 2.1.

Możemy zauważyć, że wzór 2.4 można zapisać przy założeniu, ze wszystkie zmienne X_i mają takie same rozkłady, czyli $F = F_i$, $p = p_i$ dla $i = 1, \ldots, n$. Wtedy:

$$\sum_{s=0}^{n-k} \sum_{(j_1,\ldots,j_n)} \frac{1}{z_1! \cdot \ldots \cdot z_{r-1}!(z_r+s)!(n-k-s)} \prod_{i=1}^{k} p_{j_i}(t_i) \prod_{i=k+1}^{k+s} p_{j_i}(t_k) \prod_{i=k+s+1}^{n} [1-F_{j_i}(t_k)] =$$

$$= \sum_{s=0}^{n-k} \sum_{(j_1,\ldots,j_n)} \frac{1}{z_1! \cdot \ldots \cdot z_{r-1}!(z_r+s)!(n-k-s)} \prod_{i=1}^{k} p(t_i) \prod_{i=k+1}^{k+s} p(t_k) \prod_{i=k+s+1}^{n} [1-F(t_k)] =$$

$$= \sum_{s=0}^{n-k} \frac{1}{z_1! \cdot \ldots \cdot z_{r-1}!(z_r+s)!(n-k-s)} \prod_{i=1}^{k} p(t_i) \prod_{i=k+1}^{k+s} p(t_k) \prod_{i=k+s+1}^{n} [1-F(t_k)] \underbrace{\sum_{(j_1,\ldots,j_n)} 1}_{=n!} =$$

$$= \sum_{s=0}^{n-k} \frac{n!}{(n-k-s)!(z_r+s)! \prod\limits_{j=1}^{r-1} z_j!} [\prod_{i=1}^{k} p(t_i)][p(t_k)]^s [1-F(t_k)]^{n-k-s}.$$

Zgodnie z oczekiwaniami otrzymaliśmy wzór 2.4 z twierdzenia 2.3.

UWAGA 2.2.

Podobnie, wzór 2.4 można zapisać przy założeniach twierdzenia 2.2. Zakładamy zatem, że $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi oraz $X_1 \sim G, g$ i $X_2, \ldots, X_n \overset{\text{iid}}{\sim} F, p$.

Chcemy przekształcić odpowiednio wyrażenie:

$$\sum_{s=0}^{n-k} \frac{1}{z_1! \cdot \ldots \cdot z_{r-1}!(z_r+s)!(n-k-s)} \underbrace{\sum_{(j_1,\ldots,j_n)} \prod_{i=1}^{k} p_{j_i}(t_i) \prod_{i=k+1}^{k+s} p_{j_i}(t_k) \prod_{i=k+s+1}^{n} [1-F_{j_i}(t_k)]}_{\substack{\text{rozbijemy na } n \text{ sum, w zależności od tego,} \\ \text{które z } j_i \text{ jest równe } 1}} = (\clubsuit).$$

Przekształcamy zatem zaznaczony powyżej fragment wyrażenia.

$$\underbrace{\sum_{(j_1,\ldots,j_n)} \prod_{i=1}^{k} p_{j_i}(t_i) \prod_{i=k+1}^{k+s} p_{j_i}(t_k) \prod_{i=k+s+1}^{n} [1 - F_{j_i}(t_k)]}_{\text{dla krótszego zapisu oznaczymy: } A} =$$

$$= \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_1=1}} A + \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_2=1}} A + \ldots + \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_n=1}} A =$$

$$= \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_1=1}} A + \ldots + \underbrace{\sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{z_1+\ldots+z_{m-1}+1}=1}} A + \ldots + \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{z_1+\ldots+z_m}=1}}}_{(1m*)} A + \ldots +$$

$$+ \underbrace{\sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k-z_r+1}=1}} A + \ldots + \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k+s}=1}} A}_{(2*)} + \ldots + \underbrace{\sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k+s+1}=1}} A + \ldots + \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_n=1}} A}_{(3*)} = (*).$$

Rozpiszemy teraz poszczególne fragmenty: $(1m*)$ dla $m = 1, \ldots, r-1$, $(2*)$, $(3*)$ zakładając, że $X_1 \sim G, g$ i $X_2, \ldots, X_n \overset{\text{iid}}{\sim} F, p$.

Najpierw rozważymy pierwszy fragment, czyli sumę po permutacjach zbioru $\{1, \ldots, n\}$, w których na miejscach od $z_1 + \ldots + z_{m-1} + 1$ do $z_1 + \ldots + z_m$ znajduje się 1.

$$(1m*) = \sum_{\substack{(j_1,\ldots,j_n),\ \text{gdzie}\\ j_{z_1+\ldots+z_{m-1}+1}=1}} \frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} + \ldots +$$

$$+ \sum_{\substack{(j_1,\ldots,j_n),\ \text{gdzie}\\ j_{z_1+\ldots+z_m}=1}} \frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} =$$

$$= z_m(n-1)! \frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s}.$$

Teraz popatrzymy na sumę po permutacjach, w których na miejscach od $k - z_r + 1$ do $k + s$ znajduje się 1.

$$(2*) = \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k-z_r+1}=1}} \frac{g(t_k)}{p(t_k)} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} + \ldots + \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_k=1}} \frac{g(t_k)}{p(t_k)} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s} +$$

$$+ \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k+1}=1}} + \prod_{i=1}^{k} p(t_i)[p(t_k)]^{s-1} g(t_k)[1 - F(t_k)]^{n-k-s} \quad \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k+s}=1}} \prod_{i=1}^{k} p(t_i)[p(t_k)]^{s-1} g(t_k)[1 - F(t_k)]^{n-k-s} =$$

$$= (z_r + s)(n-1)! \frac{g(t_k)}{p(t_k)} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s}.$$

Na koniec zajmiemy się sumą po permutacjach, w których na miejscach od $k + s + 1$ do n znajduje się 1.

$$(3*) = \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_{k+s+1}=1}} \prod_{i=1}^{k} (t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s-1}[1 - G(t_k)] + \ldots +$$

$$+ \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_n=1}} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s-1}[1 - G(t_k)] =$$

$$= (n - k - s)(n-1)! \frac{1 - G(t_k)}{1 - F(t_k)} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1 - F(t_k)]^{n-k-s}.$$

Zatem:

$$(*) = \sum_{m=1}^{r-1} z_m(n-1)! \frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1-F(t_k)]^{n-k-s}+$$

$$+(z_r+s)(n-1)! \frac{g(t_k)}{p(t_k)} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1-F(t_k)]^{n-k-s}+$$

$$+(n-k-s)(n-1)! \frac{1-G(t_k)}{1-F(t_k)} \prod_{i=1}^{k} p(t_i)[p(t_k)]^s[1-F(t_k)]^{n-k-s}.$$

Ostatecznie otrzymujemy tezę twierdzenia 2.2:

$$(\clubsuit) = \frac{(n-1)!}{\prod_{j=1}^{r-1} z_j!} \prod_{i=1}^{k} p(t_i) \left[\sum_{s=0}^{n-k} \frac{[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(z_r+s)!(n-k-s)} \left(\sum_{m=1}^{r-1} \frac{z_m g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} + \frac{(z_r+s)g(t_k)}{p(t_k)} + \right.\right.$$

$$\left.\left. + \frac{(n-k-s)(1-G(t_k))}{1-F(t_k)} \right) \right].$$

Znaleźliśmy zatem rozkład łączny k pierwszych statystyk pozycyjnych w trzech sytuacjach: gdy zmienne na podstawie których tworzymy statystyki mają jednakowy rozkład, gdy jedna ze zmiennych ma inny rozkład niż pozostałe oraz gdy każda ze zmiennych może mieć inny rozkład. Następnie sprowadziliśmy przypadek najbardziej ogólny do dwóch mniej ogólnych. Twierdzenie 2.1 z tego podrozdziału będzie wykorzystywane jeszcze później w przykładach w rozdziale ??, a kolejny podrozdział jest poświęcony kolejnej, bardziej ogólnej sytuacji.

2.3 Rozkład łączny: $\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q)$

Kolejną rozważaną sytuacją będzie znalezienie rozkładu łącznego: $\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q)$ dla $k < q$. Szukamy zatem prawdopodobieństwa, że pierwsze k statystyk porządkowych ma pewne określone wartości i dodatkowo pewna dalsza, q-ta statystyka ma również jakąś określoną wartość. Należy rozważyć dwa przypadki:

- gdy $t_k = t_q$, wtedy postać rozkładu $\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q)$ sprowadza się łatwo do przypadku z twierdzenia 2.1,
- gdy $t_k < t_q$.

2.3.1 Przypadek zmiennych niezależnych o jednakowym rozkładzie

Tak jak w poprzednim podrozdziale zaczniemy od najprostszej sytuacji, czyli zmiennych o jednakowym rozkładzie. Obie powyższe sytuacje zostaną opisane przy pomocy poniższego twierdzenia.

TWIERDZENIE 2.4.
Załóżmy, że $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$. *Oznaczmy* $\mathbf{X_{(k)}} = (X_{(1)}, \ldots, X_{(k)})$ *oraz* $\mathbf{t_k} = (t_1, \ldots, t_k)$.
Wtedy:

$$\mathbb{P}(\mathbf{X_{(k)}} = \mathbf{t_k}, X_{(q)} = t_q) = \begin{cases} \sum_{s=0}^{n-q} \frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r+q-k+s)!(n-q-s)!} \prod_{j=1}^{q} p(t_j)[p(t_k)]^s[1-F(t_k)]^{n-q-s} & \text{gdy } t_k = t_q, \\ \sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-1-u} \sum_{v=0}^{n-q} \frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r+u)!h!(n-k-u-h-v)!v!} \prod_{j=1}^{k} p(t_j)[p(t_k)]^u \times & \\ \times [F(t_q^-) - F(t_k)]^h [p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v & \text{gdy } t_k < t_q. \end{cases}$$

Dowód:

Spojrzymy najpierw na prostszy przypadek: $t_k = t_q$. Zakładamy, że oznaczenia dotyczące obserwacji związanych nie zmieniły się mimo, że obserwacja t_k powtarza się więcej niż z_r razy. Zatem wartości statystyk porządkowych $X_{(1)}, \ldots, X_{(q)}$ wyglądają następująco:

$$\underbrace{t_1 = \ldots = t_{z_1}}_{z_1} < \ldots < \underbrace{t_{z_1+\ldots z_{r-1}+1} = \ldots = t_{z_1+\ldots+z_r} = t_{k+1} = \ldots = t_q}_{z_r+q-k}.$$

Szukane prawdopodobieństwo wynosi:

$$\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q) = \mathbb{P}(X_{(1)} = t_1, \ldots, X_{(q)} = t_k) = (*),$$

zatem korzystając z twierdzenia 2.1:

$$(*) = \sum_{s=0}^{n-q} \frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r + q - k + s)!(n - q - s)!} \prod_{j=1}^{q} p(t_j)[p(t_k)]^s[1 - F(t_k)]^{n-q-s}.$$

Drugim przypadkiem jest sytuacja: $t_k < t_q$. Oznacza to, że mogą istnieć obserwacje leżące pomiędzy t_k i t_q. Sytuację tę można zilustrować przy pomocy poniższego rysunku.

$$\underbrace{\overbrace{X_{(1)} \cdots X_{(z_1)}}^{z_1}}_{=t_{z_1}} \cdots \underbrace{\overbrace{X_{(k-z_r+1)} \cdots X_{(k+u)}}^{z_r+u}}_{=t_k} \underbrace{\overbrace{X_{(k+u+1)} \cdots X_{k+u+h}}^{h}}_{\in(t_k,t_q)} \underbrace{\overbrace{X_{(k+u+h+1)} \cdots X_{(n-v)}}^{n-k-u-h-v}}_{=t_q} \underbrace{\overbrace{X_{(n-v+1)} \cdots X_{(n)}}^{v}}_{>t_q}$$

dla:

$u = 0, \ldots, q - k - 1,$

$h = 0, \ldots, q - k - u - 1,$

$v = 0, \ldots, n - q.$

Zdefiniujmy zatem dla powyższych u, h, v zdarzenia:

$$A_{uhv} : \begin{cases} z_1 & \text{obserwacji jest równych } t_1 = \ldots = t_{z_1} & \text{ozn. } B_1, \\ z_2 & \text{obserwacji jest równych } t_{z_1+1} = \ldots = t_{z_1+z_2} & \text{ozn. } B_2, \\ \vdots & & \\ z_r + u & \text{obserwacji jest równych } t_{z_1+\ldots+z_r} = t_k & \text{ozn. } B_{(u)r}, \\ h & \text{obserwacji należy do przedziału } (t_k, t_q) & \text{ozn. } C_{(h)}, \\ n - k - u - h - v & \text{obserwacji jest równych } t_q & \text{ozn. } D_{(uhv)}, \\ v & \text{obserwacji jest większych od } t_q & \text{ozn. } E_{(u)r}. \end{cases}$$

Zatem szukane prawdopodobieństwo wynosi:

$$\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q) = \mathbb{P}\Big(\bigcup_{u=0}^{q-k-1} \bigcup_{h=0}^{q-k-u-1} \bigcup_{v=0}^{n-q} A_{uhv}\Big) = (*),$$

a ponieważ zdarzenia A_{uhv} są parami rozłączne, to:

$$(*) = \sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-u-1} \sum_{v=0}^{n-q} \mathbb{P}(A_{uhv}) = (**).$$

Teraz trzeba znaleźć $\mathbb{P}(A_{uhv})$ i zrobimy to analogicznie do wyprowadzenia z dowodu twierdzenia 2.1.

$$
\begin{aligned}
\mathbb{P}(A_{uhv}) &= \mathbb{P}(B_1)\mathbb{P}(B_2|B_1)\cdot\ldots\cdot\mathbb{P}(B_{(u)r}|B_1,\ldots,B_{r-1})\mathbb{P}(C_{(h)}|B_1,\ldots,B_{(u)r})\times \\
&\quad\times \mathbb{P}(D_{(uhv)}|B_1,\ldots,B_{(u)r},C_{(h)})\mathbb{P}(E_{(v)}|B_1,\ldots,B_{(u)r},C_{(h)},D_{(uhv)}) = \\
&= \binom{n}{z_1}[p(t_{z_1})]^{z_1}\cdot\ldots\cdot\binom{n-z_1-\ldots-z_{r-1}}{z_r+u}[p(t_{z_1+\ldots+z_r})]^{z_r+u}\binom{n-k-u}{h}[F(t_q^-)-F(t_k)]^h\times \\
&\quad\times\binom{n-k-u-h}{n-k-u-h-v}[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v = \\
&= \frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r+u)!h!(n-k-u-h-v)!v!}[\prod_{i=1}^{k} p(t_i)][p(t_k)]^u[F(t_q^-)-F(t_k)]^h\times \\
&\qquad\qquad\times[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v.
\end{aligned}
$$

Zatem ostatecznie:

$$
\begin{aligned}
(**) = \sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-u-1}\sum_{v=0}^{n-q}&\frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r+u)!h!(n-k-u-h-v)!v!}[\prod_{i=1}^{k} p(t_i)][p(t_k)]^u\times \\
&\times[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v.
\end{aligned}
$$

$\square$

2.3.2 Przypadek zmiennych niezależnych o dowolnych rozkładach

Tym razem od razu opiszemy z pozoru trudniejszy przypadek, czyli sytuację w której każda zmienna X_i ma inny rozkład. Pomijamy sytuację, gdy jedna ze zmiennych ma inny rozkład od pozostałych.

TWIERDZENIE 2.5.

Niech $X_i \sim F_i, p_i$ dla $i = 1,\ldots,n$ oraz zmienne X_i są niezależne. Oznaczmy $\mathbb{X}_{(k)} = (X_{(1)},\ldots,X_{(k)})$ oraz $\mathbb{t}_k = (t_1,\ldots,t_k)$.

Wtedy:

$$
\mathbb{P}(\mathbb{X}_{(k)}=\mathbb{t}_k, X_{(q)}=t_q) = \begin{cases}
\sum\limits_{s=0}^{n-q}\sum\limits_{\substack{(j_1,\ldots,j_n)-\\ \textit{permutacje zb.}\\ \{1,\ldots,n\}}}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+q-k+s)!(n-q-s)!}\prod\limits_{i=1}^{q} p_{j_i}(t_i)\prod\limits_{i=q+1}^{q+s} p_{j_i}(t_k)\times \\
\qquad\qquad\times\prod\limits_{i=q+s+1}^{n}[1-F_{j_i}(t_k)] & \textit{gdy } t_k = t_q, \\
\sum\limits_{u=0}^{q-k-1}\sum\limits_{h=0}^{q-k-u-1}\sum\limits_{v=0}^{n-q}\sum\limits_{\substack{(j_1,\ldots,j_n)-\\ \textit{permutacje zb.}\\ \{1,\ldots,n\}}}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!}\times \\
\qquad\times\prod\limits_{i=1}^{k} p_{j_i}(t_i)\prod\limits_{i=k+1}^{k+s} p_{j_i}(t_k)\prod\limits_{k+u+1}^{k+u+h}[F_{j_i}(t_q^-)-F_{j_i}(t_k)]\times \\
\qquad\times\prod\limits_{i=k+u+h+1}^{n-v} p_{j_i}(t_q)\prod\limits_{i=n-v+1}^{n}[1-F_{j_i}(t_q)] & \textit{gdy } t_k < t_q,
\end{cases} =
$$

$$
= \begin{cases}
\sum\limits_{s=0}^{n-q}\sum\limits_{\substack{(j_1,\ldots,j_n)-\\ \textit{permutacje zb.}\\ \{1,\ldots,n\}}}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+q-k+s)!(n-q-s)!}PerM_s^{(1)} & \textit{gdy } t_k = t_q, \\
\sum\limits_{u=0}^{q-k-1}\sum\limits_{h=0}^{q-k-u-1}\sum\limits_{v=0}^{n-q}\sum\limits_{\substack{(j_1,\ldots,j_n)-\\ \textit{permutacje zb.}\\ \{1,\ldots,n\}}}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!}PerM_{uhv}^{(2)} & \textit{gdy } t_k < t_q.
\end{cases}
$$

gdzie:

$$M^{(2)}_{uhv} = \begin{bmatrix} p_1(t_1) & \dots & p_n(t_1) \\ p_1(t_2) & \dots & p_n(t_2) \\ \vdots & & \vdots \\ p_1(t_k) & \dots & p_n(t_k) \\ \vdots & & \vdots \\ p_1(t_k) & \dots & p_n(t_k) \\ F_1(t_q^-) - F_1(t_k) & \dots & F_n(t_q^-) - F_n(t_k) \\ \vdots & & \vdots \\ F_1(t_q^-) - F_1(t_k) & \dots & F_n(t_q^-) - F_n(t_k) \\ p_1(t_q) & \dots & p_n(t_q) \\ \vdots & & \vdots \\ p_1(t_q) & \dots & p_n(t_q) \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \\ \vdots & & \vdots \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a\\a\\a\\a\\a\\a \end{matrix}}\right\} k+u \\ \left.\vphantom{\begin{matrix} a\\a\\a \end{matrix}}\right\} h \\ \left.\vphantom{\begin{matrix} a\\a\\a \end{matrix}}\right\} n-k-u-h-v \\ \left.\vphantom{\begin{matrix} a\\a\\a \end{matrix}}\right\} v \end{matrix} \qquad M^{(1)}_s = \begin{bmatrix} p_1(t_1) & \dots & p_n(t_1) \\ p_1(t_2) & \dots & p_n(t_2) \\ \vdots & & \vdots \\ p_1(t_q) & \dots & p_n(t_q) \\ \vdots & & \vdots \\ p_1(t_q) & \dots & p_n(t_q) \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \\ \vdots & & \vdots \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a\\a\\a\\a\\a\\a \end{matrix}}\right\} q+s \\ \left.\vphantom{\begin{matrix} a\\a\\a \end{matrix}}\right\} n-q-s \end{matrix}$$

Dowód:

Ponownie musimy rozpatrzeć dwa przypadki: $t_k = t_q$ oraz $t_k < t_q$.

Pierwszy z nich można łatwo sprowadzić do sytuacji opisanej w twierdzeniu 2.3, uwzględniając jedynie zmiany: $k \to q$ oraz $z_r \to z_r + q - k$, ponieważ wtedy szukane prawdopodobieństwo jest równe:

$$\mathbb{P}(X_{(1)} = t_1, \dots, X_{(k)} = t_k, X_{(q)} = t_q) = \mathbb{P}(X_{(1)} = t_1, \dots, X_{(q)} = t_k) = (*),$$

a r-ty blok związanych obserwacji jest dłuższy o $q - k$. Zatem:

$$(*) = \sum_{s=0}^{n-q} \sum_{\substack{(j_1,\dots,j_n)-\\ \text{permutacje zb.}\\ \{1,\dots,n\}}} \frac{1}{\prod_{i=1}^{r-1} z_i!(z_r + q - k + s)!(n-q-s)!} \prod_{i=1}^{q} p_{j_i}(t_i) \prod_{i=q+1}^{q+s} p_{j_i}(t_i) \prod_{i=q+s+1}^{n} [1 - F_{j_i}(t_k)] = (**).$$

Jeżeli dodatkowo dla $s = 1, \dots, n-k$ zdefiniujemy macierze:

$$M^{(1)}_s = \begin{bmatrix} p_1(t_1) & \dots & p_n(t_1) \\ p_1(t_2) & \dots & p_n(t_2) \\ \vdots & & \vdots \\ p_1(t_q) & \dots & p_n(t_q) \\ \vdots & & \vdots \\ p_1(t_q) & \dots & p_n(t_q) \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \\ \vdots & & \vdots \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a\\a\\a\\a\\a\\a \end{matrix}}\right\} q+s \\ \left.\vphantom{\begin{matrix} a\\a\\a \end{matrix}}\right\} n-q-s \end{matrix}$$

to powyższy wzór możemy przy pomocy definicji permanentu (definicja 1.3) zapisać następująco:

$$(**) = \sum_{s=0}^{n-q} \sum_{\substack{(j_1,\dots,j_n)-\\ \text{permutacje zb.}\\ \{1,\dots,n\}}} \frac{1}{\prod_{i=1}^{r-1} z_i!(z_r + q - k + s)!(n-q-s)!} \mathrm{Per} M^{(1)}_s.$$

Dowód twierdzenia w drugim przypadku będzie podobny do dowodów twierdzeń 2.3 oraz 2.4. Najpierw zdefiniujemy parami rozłączne zdarzenia A_{uvh} tak jak w dowodzie twierdzenia 2.4:

$$A_{uhv}: \begin{cases} z_1 & \text{obserwacji jest równych } t_1 = \ldots = t_{z_1}, \\ z_2 & \text{obserwacji jest równych } t_{z_1+1} = \ldots = t_{z_1+z_2}, \\ \vdots & \\ z_r + u & \text{obserwacji jest równych } t_{z_1+\ldots+z_r} = t_k, \\ h & \text{obserwacji należy do przedziału } (t_k, t_q), \\ n-k-u-h-v & \text{obserwacji jest równych } t_q, \\ v & \text{obserwacji jest większych od } t_q, \end{cases}$$

dla
$u = 0, \ldots, q-k-1,$
$h = 0, \ldots, q-k-u-1,$
$v = 0, \ldots, n-q.$
Wtedy:

$$\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q) = \mathbb{P}(\bigcup_{u=0}^{q-k-1} \bigcup_{h=0}^{q-k-u-1} \bigcup_{v=0}^{n-q} A_{uhv}) = \sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-u-1} \sum_{v=0}^{n-q} \mathbb{P}(A_{uhv}) = (**).$$

Aby znaleźć prawdopodobieństwo $\mathbb{P}(A_{uhv})$ zdefiniujmy zdarzenia:

$$P_{(j_1,\ldots,j_n)}: \begin{cases} X_{j_1} = \ldots = X_{j_{z_1}} = t_{z_1}, \\ \vdots \\ X_{j_{z_1+\ldots+z_{r-1}+1}} = \ldots = X_{j_{k+u}} = t_k, \\ X_{j_{k+u+1}}, \ldots, X_{j_{k+u+h}} \in (t_k, t_q), \\ X_{j_{k+u+h+1}} = \ldots = X_{j_{n-v}} = t_q, \\ X_{j_{n-v+1}}, \ldots, X_{j_n} > t_q. \end{cases}$$
gdzie $(j_1, \ldots, j_n)$ - dowolna permutacja $\{1, \ldots, n\}$.

Wtedy:

$$(**) = \sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-u-1} \sum_{v=0}^{n-q} \mathbb{P}(\bigcup_{\substack{(j_1,\ldots,j_n)-\\ \text{permutacje zb.}\\ \{1,\ldots,n\}}} P_{(j_1,\ldots,j_n)}) = (***).$$

Trzeba zliczyć ile razy każde zdarzenie występuje w sumie $\bigcup_{(j_1,\ldots,j_n)} P_{(j_1,\ldots,j_n)}$. Zliczymy je tak samo jak w dowodzie twierdzenia 2.3:

$$\underbrace{X_{j_1} \ldots X_{j_{z_1}}}_{\substack{z_1!\\ \text{permutacji}}} \ldots \underbrace{X_{j_{z_1+\ldots+z_{r-1}+1}} \ldots X_{j_k} \ldots X_{j_{k+u}}}_{\substack{(z_r+u)!\\ \text{permutacji}}} \underbrace{X_{j_{k+u+1}} \ldots X_{j_{k+u+h}}}_{\substack{h!\\ \text{permutacji}}} \underbrace{X_{j_{k+u+h+1}} \ldots X_{j_{n-v}}}_{\substack{(n-k-u-h-v)!\\ \text{permutacji}}} \underbrace{X_{j_{n-v+1}} \ldots X_{j_n}}_{\substack{v!\\ \text{permutacji}}}$$

Zatem każde ze zdarzeń powtarza się $\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!$ razy, więc:

$$\mathbb{P}(\bigcup_{\substack{(j_1,\ldots,j_n)-\\ \text{permutacje zb.}\\ \{1,\ldots,n\}}} P_{(j_1,\ldots,j_n)}) = \sum_{\substack{(j_1,\ldots,j_n)-\\ \text{permutacje zb.}\\ \{1,\ldots,n\}}} \frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!} \mathbb{P}(P_{(j_1,\ldots,j_n)}).$$

Ze względu na niezależność obserwacji $X_1, \ldots, X_n$, zachodzi:

$$\mathbb{P}(P_{(j_1,\ldots,j_n)}) = \prod_{i=1}^{k} p_{j_i}(t_i) \prod_{i=k+1}^{k+u} p_{j_i}(t_i) \prod_{k+u+1}^{k+u+h} [F_{j_i}(t_q^-) - F_{j_i}(t_k)] \prod_{i=k+u+h+1}^{n-v} p_{j_i}(t_q) \prod_{i=n-v+1}^{n} [1 - F_{j_i}(t_q)],$$

zatem ostatecznie

$$(***) = \sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-u-1}\sum_{v=0}^{n-q}\sum_{\substack{(j_1,\dots,j_n)-\\ \text{permutacje zb.}\\ \{1,\dots,n\}}} \frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!} \times$$

$$\times \prod_{i=1}^{k} p_{j_i}(t_i) \prod_{i=k+1}^{k+u} p_{j_i}(t_i) \prod_{k+u+1}^{k+u+h} [F_{j_i}(t_q^-) - F_{j_i}(t_k)] \prod_{i=k+u+h+1}^{n-v} p_{j_i}(t_q) \prod_{i=n-v+1}^{n} [1 - F_{j_i}(t_q)] = (\clubsuit).$$

Jeżeli dodatkowo dla odpowiednich u, h, v zdefiniujemy macierze:

$$M_{uhv}^{(2)} = \begin{bmatrix} p_1(t_1) & \dots & p_n(t_1) \\ p_1(t_2) & \dots & p_n(t_2) \\ \vdots & & \vdots \\ p_1(t_k) & \dots & p_n(t_k) \\ \vdots & & \vdots \\ p_1(t_k) & \dots & p_n(t_k) \\ F_1(t_q^-) - F_1(t_k) & \dots & F_n(t_q^-) - F_n(t_k) \\ \vdots & & \vdots \\ F_1(t_q^-) - F_1(t_k) & \dots & F_n(t_q^-) - F_n(t_k) \\ p_1(t_q) & \dots & p_n(t_q) \\ \vdots & & \vdots \\ p_1(t_q) & \dots & p_n(t_q) \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \\ \vdots & & \vdots \\ 1 - F_1(t_q) & \dots & 1 - F_n(t_q) \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} a \\ a \\ \vdots \\ a \\ \vdots \\ a \end{matrix}}\right\} k+u \\ \left.\vphantom{\begin{matrix} a \\ \vdots \\ a \end{matrix}}\right\} h \\ \left.\vphantom{\begin{matrix} a \\ \vdots \\ a \end{matrix}}\right\} n-k-u-h-v \\ \left.\vphantom{\begin{matrix} a \\ \vdots \\ a \end{matrix}}\right\} v \end{array}$$

to powyższy wzór przyjmie postać:

$$(\clubsuit) = \sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-u-1}\sum_{v=0}^{n-q}\sum_{\substack{(j_1,\dots,j_n)-\\ \text{permutacje zb.}\\ \{1,\dots,n\}}} \frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!} \operatorname{Per} M_{uhv}^{(2)}.$$

□

UWAGA 2.3.

Podobnie jak w uwadze 2.1 do twierdzenia 2.3, we wzorze z powyższego twierdzenia 2.5 można dodatkowo założyć, że zmienne $X_1, \dots, X_n$ mają jednakowe rozkłady i wtedy otrzymamy wzór z twierdzenia 2.4.

UWAGA 2.4.

W tym podrozdziale rozważyliśmy sytuację zmiennych niezależnych $X_1, \dots, X_n$ o jednakowym rozkładzie, a następnie przeszliśmy od razu do przypadku, gdy zmienne te maja dowolne rozkłady. Pominęliśmy przypadek, gdy tylko jedna zmienna X_1 ma inny rozkład. Wynikało to z faktu, że rachunki byłyby analogiczne do tych z twierdzenia 2.2 (mówiącego o rozkładzie k pierwszych statystyk porządkowych w przypadku gdy jedna zmienna ma inny rozkład), a jedynie bardziej złożone. Teraz jednak możemy, korzystając z udowodnionego twierdzenia 2.5, założyć, że $X_1 \sim G, g$ i jest niezależna od $X_2, \dots, X_n \overset{\text{iid}}{\sim} F, p$ i wtedy przekształcić wzór z twierdzenia 2.5. Zrobimy to tylko w sytuacji, gdy $t_k < t_q$.

Zgodnie z oznaczeniami w twierdzeniu 2.5, w ogólnej postaci wzór przyjmuje postać:

$$\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-u-1}\sum_{v=0}^{n-q}\frac{1}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!}\underbrace{\sum_{\substack{(j_1,\ldots,j_n)-\\ \text{permutacje zb.}\\ \{1,\ldots,n\}}}\mathrm{Per}M^{(2)}_{uhv}}_{(*)} = (\clubsuit).$$

Rozpiszemy zaznaczony fragment:

$$(*) = \sum_{\substack{(j_1,\ldots,j_n)-\\ \text{permutacje zb.}\\ \{1,\ldots,n\}}}\underbrace{\mathrm{Per}M^{(2)}_{uhv}}_{oznaczmy: A} =$$

$$= \sum_{\substack{(j_1,\ldots,j_n)\\ \text{gdzie } j_1=1}} A+\ldots+\underbrace{\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{z_1+\ldots+z_{m-1}+1}=1}} A+\ldots+\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{z_1+\ldots+z_m}=1}} A}_{(1m*)}+\ldots+$$

$$+\underbrace{\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{k-z_r+1}=1}} A+\ldots+\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{k+u}=1}} A}_{(1r*)}+\underbrace{\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{k+u+1}=1}} A+\ldots+\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{k+u+h}=1}} A}_{(2*)}+$$

$$+\underbrace{\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{k+u+h+1}=1}} A+\ldots+\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{n-v}=1}} A}_{(3*)}+\underbrace{\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_{n-v+1}=1}} A+\ldots+\sum_{\substack{(j_1,\ldots,j_n),\text{ gdzie}\\ j_n=1}} A}_{(4*)} = (**).$$

Rozpisując każdy z elementów powyższej sumy analogicznie jak w uwadze 2.2 otrzymamy:

$$(1m*) = z_m(n-1)!\frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})}\prod_{i=1}^{k}p(t_i)[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-v-u-h}[1-F(t_q)]^v,$$

$$(1r*) = (z_r+u)(n-1)!\frac{g(t_k)}{p(t_k)}\prod_{i=1}^{k}p(t_i)[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-v-u-h}[1-F(t_q)]^v,$$

$$(2*) = h(n-1)!\frac{G(t_q^-)-G(t_k)}{F(t_q^-)-F(t_k)}\prod_{i=1}^{k}p(t_i)[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-v-u-h}[1-F(t_q)]^v,$$

$$(3*) = (n-k-h-u-v)(n-1)!\frac{g(t_q)}{p(t_q)}\prod_{i=1}^{k}p(t_i)[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-v-u-h}[1-F(t_q)]^v,$$

$$(4*) = v(n-1)!\frac{1-G(t_q)}{1-F(t_q)}\prod_{i=1}^{k}[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-v-u-h}[1-F(t_q)]^v.$$

Zatem:

$$(**) = (n-1)!\prod_{i=1}^{k}p(t_i)[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-v-u-h}[1-F(t_q)]^v\left[z_m\frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})}+\right.$$

$$\left.+(z_r+u)\frac{g(t_k)}{p(t_k)}+h\frac{G(t_q^-)-G(t_k)}{F(t_q^-)-F(t_k)}+(n-k-h-u-v)\frac{g(t_q)}{p(t_q)}+v\frac{1-G(t_q)}{1-F(t_q)}\right].$$

Ostatecznie otrzymujemy:

$$(\clubsuit) = \sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-u-1} \sum_{v=0}^{n-q} \frac{(n-1)!}{\prod_{i=1}^{r-1} z_i!(z_r+u)!h!(n-k-u-h-v)!v!} \times$$

$$\times \prod_{i=1}^{k} p(t_i)[p(t_k)]^u [F(t_q^-) - F(t_k)]^h [p(t_q)]^{n-k-v-u-h} [1-F(t_q)]^v \times$$

$$\times \left[z_m \frac{g(t_{z_1+\ldots+z_m})}{p(t_{z_1+\ldots+z_m})} + (z_r+u)\frac{g(t_k)}{p(t_k)} + h\frac{G(t_q^-) - G(t_k)}{F(t_q^-) - F(t_k)} + (n-k-h-u-v)\frac{g(t_q)}{p(t_q)} + v\frac{1-G(t_q)}{1-F(t_q)} \right].$$

Wyprowadziliśmy zatem wzór na rozkład łączny w sytuacji, gdy jedna ze zmiennych na inny rozkład od pozostałych, czyli w próbie z jedną obserwacją odstającą.

2.4 Rozkład łączny dowolnego podciągu: $\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k})$

Ostatnim rozkładem łącznym jaki znajdziemy będzie rozkład łączny dowolnego podciągu statystyk porządkowych. Jest to przypadek najbardziej ogólny i zawiera te omówione w podrozdziałach 2.2 i 2.3. Chcemy znaleźć rozkład łączny dowolnego podciągu indeksów $(i_1, \ldots, i_k) \subset (1, \ldots, n)$, zatem będziemy szukać:

$$\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}).$$

Wykorzystamy w tym celu notację przyjętą w definicji 1.2 związanych bloków obserwacji, czyli:

$$t_{i_1} = t_{i_2} = \ldots = t_{i_{z_1}} < t_{i_{z_1}+1} = \ldots = t_{i_{z_1+z_2}} < \ldots < t_{i_{z_1+\ldots+z_{r-1}}+1} = \ldots = t_{i_{z_1+\ldots+z_r}} = t_{i_k}.$$

2.4.1 Przypadek zmiennych niezależnych o jednakowym rozkładzie

Ponownie na początku rozważymy przypadek, gdy $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$.

TWIERDZENIE 2.6.

Niech $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$ oraz $t_{i_1}, \ldots, t_{i_k}$ ma r związanych bloków. Przyjmijmy też dodatkowo, że $v_0 = 0$, $v_{r+1} = n$, $s_{r+1} = n - v_r$, $t_{i_{z_1+\ldots+z_0}} = -\infty$, $t_{i_{z_1+\ldots+z_{r+1}}} = \infty$.

Wtedy:

$$\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}) = \sum_{s_1=0}^{i_1-1} \sum_{v_1=i_{z_1}}^{i_{z_1+1}-1} \sum_{s_2=0}^{i_{z_1+1}-1-v_1} \sum_{v_2=i_{z_1+z_2}}^{i_{z_1+z_2+1}-1} \cdot \ldots \cdot \sum_{s_r=0}^{i_{k-z_r+1}-1-v_{r-1}} \times$$

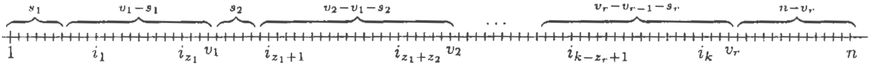
$$\times \sum_{v_r=i_k}^{n} \prod_{j=1}^{r+1} \frac{n!}{s_j!(v_j - v_{j-1} - s_j)!} [F(t^-_{i_{z_1+\ldots+z_j}}) - F(t_{i_{z_1+\ldots+z_{j-1}}})]^{s_j} [p(t_{i_{z_1+\ldots+z_j}})]^{v_j - v_{j-1} - s_j}.$$

Dowód:

Dowód jest analogiczny do dowodów twierdzeń 2.1 i 2.4, jednak z uwagi na ogólniejszy charakter tego twierdzenia, będzie on wymagać bardziej złożonej notacji.

Zauważmy najpierw, że zdarzenie „$X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}$" oznacza, że pewna część obserwacji jest równa $t_{i_{z_1+\ldots+z_j}}$ dla różnych j, a pozostałe są albo mniejsze od t_{i_1}, albo większe od t_{i_k}, albo leżą pomiędzy $t_{i_{z_1+\ldots+z_{j-1}}}$, a $t_{i_{z_1+\ldots+z_j}}$ dla różnych j.

Opisaną sytuację najlepiej zobrazować poniższym rysunkiem, w którym dla większej czytelności zaznaczone są jedynie indeksy statystyk porządkowych.



Dodatkowe wartości $s_1, \ldots, s_r$ oznaczają odpowiednio liczbę obserwacji, które nie są równe $t_{z_1+\ldots+z_j}$ dla żadnego j, lecz leżą gdzieś pomiędzy tymi wartościami. Natomiast $v_1, \ldots, v_r$ są punktami na osi indeksów. Tak zdefiniowane $s_1, \ldots, s_r$ i $v_1, \ldots, v_r$ mogą przyjmować następujące wartości.

$$\begin{cases} s_1 = 0, \ldots, i_1 - 1, & v_1 = i_{z_1}, \ldots, i_{z_1+1} - 1, \\ \vdots & \vdots \\ s_k = 0, \ldots, i_{z_1+\ldots+z_{k-1}+1} - 1 - v_{k-1}, & v_k = i_{z_1+\ldots+z_k}, \ldots, i_{z_1+\ldots+z_k+1} - 1, \\ \vdots & \vdots \\ s_r = 0, \ldots, i_{z_1+\ldots+z_{r-1}+1} - 1 - v_{r-1} = & v_r = i_{z_1+\ldots+z_r}, \ldots, n = \\ \quad = 0, \ldots, i_{k-z_r+1} - 1 - v_{r-1}, & \quad = i_k, \ldots, n. \end{cases}$$

Zdefiniujmy dla powyższych $\mathbf{s} = (s_1 \ldots, s_r)$ i $\mathbf{v} = (v_1, \ldots, v_r)$ następujące zdarzenia:

$$A_{\mathbf{s},\mathbf{v}} : \begin{cases} s_1 & \text{obserwacji jest mniejszych od } t_{i_{z_1}} & (\text{ozn. } B_{1,(\mathbf{s},\mathbf{v})}), \\ v_1 - s_1 & \text{obserwacji jest równych } t_{i_{z_1}} & (\text{ozn. } C_{1(\mathbf{s},\mathbf{v})}), \\ \vdots & & \\ s_k & \text{obserwacji należy do } (t_{iz_1+\ldots+z_{k-1}}, t_{iz_1+\ldots+z_k}) & (\text{ozn. } B_{k,(\mathbf{s},\mathbf{v})}), \\ v_k - v_{k-1} - s_k & \text{obserwacji jest równych } t_{iz_1+\ldots+z_k} & (\text{ozn. } C_{k,(\mathbf{s},\mathbf{v})}), \\ \vdots & & \\ s_r & \text{obserwacji należy do } (t_{iz_1+\ldots+z_{r-1}}, t_{iz_1+\ldots+z_r}) & (\text{ozn. } B_{r,(\mathbf{s},\mathbf{v})}), \\ v_r - v_{r-1} - s_r & \text{obserwacji jest równych } t_{i_k} & (\text{ozn. } C_{r,(\mathbf{s},\mathbf{v})}), \\ n - v_r & \text{obserwacji jest większych od } t_{i_k} & (\text{ozn. } D_{(\mathbf{s},\mathbf{v})}). \end{cases}$$

Wtedy szukane prawdopodobieństwo można zapisać:

$$\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}) = \mathbb{P}(\bigcup_{\mathbf{s},\mathbf{v}} A_{\mathbf{s},\mathbf{v}}) = (*).$$

Zdarzenia $A_{\mathbf{s},\mathbf{v}}$ są parami rozłączne dla powyższych zakresów, zatem:

$$(*) = \sum_{\mathbf{s},\mathbf{v}} \mathbb{P}(A_{\mathbf{s},\mathbf{v}}) = (\clubsuit).$$

Pozostaje znalezienie prawdopodobieństwa $\mathbb{P}(A_{\mathbf{s},\mathbf{v}}) = \mathbb{P}(B_1, C_1, \ldots, B_r, C_r, D)$ w sposób analogiczny do tego z dowodów 2.1 i 2.4.

$$
\begin{aligned}
\mathbb{P}(A_{\mathbf{s}.\mathbf{v}}) &= \mathbb{P}(B_{1,(\mathbf{s},\mathbf{v})}, C_{1,(\mathbf{s},\mathbf{v})}, \ldots, B_{k,(\mathbf{s},\mathbf{v})}, C_{k,(\mathbf{s},\mathbf{v})}, \ldots, B_{r,(\mathbf{s},\mathbf{v})}, C_{r,(\mathbf{s},\mathbf{v})}, D_{(\mathbf{s},\mathbf{v})}) = \\
&= \mathbb{P}(B_{1,(\mathbf{s})})\mathbb{P}(C_{1,(\mathbf{s}}|B_{1,(\mathbf{s})}) \cdot \ldots \cdot \mathbb{P}(D_{(\mathbf{s},\mathbf{v})}|B_{1,(\mathbf{s},\mathbf{v})}, C_{1,(\mathbf{s},\mathbf{v})}, \ldots, B_{k,(\mathbf{s},\mathbf{v})}, C_{k,(\mathbf{s},\mathbf{v})}, \ldots, B_{r,(\mathbf{s},\mathbf{v})}, C_{r,(\mathbf{s},\mathbf{v})}) = \\
&= \binom{n}{s_1}[F(t^-_{i_{z_1}})]^{s_1}\binom{n-s_1}{v_1-s_1}[p(t_{i_{z_1}})]^{v_1-s_1} \cdot \ldots \cdot \binom{n-v_{k-1}}{s_k}[F(t^-_{i_{z_1+\ldots+z_{k-1}}}) - F(t_{i_{z_1+\ldots+z_k}})]^{s_k} \times \\
&\quad \times \binom{n-v_{k-1}-s_k}{v_k-v_{k-1}-s_k}[p(t_{i_{z_1+\ldots+z_k}})]^{v_k-v_{k-1}-s_k} \cdot \ldots \cdot \binom{n-v_{r-1}}{s_r}[F(t^-_{i_{z_1+\ldots+z_{r-1}}}) - F(t_{i_k})]^{s_r} \times \\
&\quad \times \binom{n-v_{r-1}-s_r}{v_r-v_{r-1}-s_r}[p(t_{i_{z_1+\ldots+z_r}})]^{v_r-v_{r-1}-s_r}\binom{n-v_r}{n-v_r}[1-F(t_{i_k})]^{n-v_r} = \\
&= \frac{n!}{s_1!(v_1-s_1)!s_2!(v_2-v_1-s_2)! \cdot \ldots \cdot s_k!(v_k-v_{k-1}-s_k)! \cdot \ldots \cdot s_r!(v_r-v_{r-1}-s_r)!(n-v_r)!} \times \\
&\quad \times [F(t^-_{i_{z_1}})]^{s_1}[F(t^-_{i_{z_1+z_2}}) - F(t_{i_{z_1}})]^{s_2} \cdot \ldots \cdot [F(t^-_{i_k}) - F(t_{i_{k-z_r}})]^{s_r}[1-F(t_{i_k})]^{n-v_r} \times \\
&\quad \times [p(t_{i_{z_1}})]^{v_1-s_1}[p(t_{i_{z_1+z_2}})]^{v_2-v_1-s_2} \cdot \ldots \cdot [p(t_{i_k})]^{v_r-v_{r-1}-s_r} = \\
&= \frac{n!}{[\prod_{j=1}^{r} s_j!](v_1-s_1)![\prod_{j=2}^{r}(v_j-v_{j-1}-s_j)!](n-v_r)!}[F(t^-_{i_{z_1}})]^{s_1}\prod_{j=2}^{r}[F(t^-_{i_{z_1+\ldots+z_j}}) - F(t_{i_{z_1+\ldots+z_{j-1}}})]^{s_j} \times \\
&\quad \times [1-F(t_{i_k})]^{n-v_r}[p(t_{i_{z_1}})]^{v_1-s_1}\prod_{j=2}^{r}[p(t_{i_{z_1+\ldots+z_j}})]^{v_j-v_{j-1}-s_j} = (**).
\end{aligned}
$$

Jeżeli dodatkowo przyjmiemy, że:

$$
\begin{cases}
v_0 = 0 & \\
v_{r+1} = n & \Rightarrow [p(t_{i_{z_1+\ldots+z_{r+1}}})]^{v_{r+1}-v_r-s_{r+1}} = [p(t_{i_{z_1+\ldots+z_{r+1}}})]^0 = 1, \\
s_{r+1} = n - v_r & \\
t_{i_{z_1+\ldots+z_0}} = -\infty & \Rightarrow [F(t^-_{i_{z_1}})]^{s_1} = [F(t^-_{i_{z_1}}) - F(t_{i_{z_1+\ldots+z_0}})]^{s_1}, \\
t_{i_{z_1+\ldots+z_{r+1}}} = \infty & \Rightarrow [1-F(t^-_{i_k})]^{n-v_r} = [F(t^-_{i_{z_1+\ldots+z_{r+1}}}) - F(t_{i_{z_1+\ldots+z_r}})]^{s_r},
\end{cases}
$$

to wtedy powyższy wzór można zapisać w postaci:

$$
(**) = \prod_{j=1}^{r+1} \frac{n!}{s_j!(v_j-v_{j-1}-s_j)!}[F(t^-_{i_{z_1+\ldots+z_j}}) - F(t_{i_{z_1+\ldots+z_{j-1}}})]^{s_j}[p(t_{i_{z_1+\ldots+z_j}})]^{v_j-v_{j-1}-s_j}.
$$

Zatem:

$$
\begin{aligned}
(\clubsuit) &= \sum_{\mathbf{s},\mathbf{v}}\prod_{j=1}^{r+1} \frac{n!}{s_j!(v_j-v_{j-1}-s_j)!}[F(t^-_{i_{z_1+\ldots+z_j}}) - F(t_{i_{z_1+\ldots+z_{j-1}}})]^{s_j}[p(t_{i_{z_1+\ldots+z_j}})]^{v_j-v_{j-1}-s_j} = \\
&= \sum_{s_1=0}^{i_1-1}\sum_{v_1=i_{z_1}}^{i_{z_1+1}-1}\sum_{s_2=0}^{i_{z_1+1}-1-v_1}\sum_{v_2=i_{z_1+z_2}}^{i_{z_1+z_2+1}-1} \cdot \ldots \cdot \sum_{s_r=0}^{i_{k-z_r+1}-1-v_{r-1}}\sum_{v_r=i_k}^{n}\prod_{j=1}^{r+1}\frac{n!}{s_j!(v_j-v_{j-1}-s_j)!} \times \\
&\quad \times [F(t^-_{i_{z_1+\ldots+z_j}}) - F(t_{i_{z_1+\ldots+z_{j-1}}})]^{s_j}[p(t_{i_{z_1+\ldots+z_j}})]^{v_j-v_{j-1}-s_j}.
\end{aligned}
$$

□

UWAGA 2.5.

Powyższy wzór można sprowadzić do wzorów z twierdzeń 2.1 i 2.4, czyli zakładając, że ciąg indeksów wynosi odpowiednio: $(i_1, \ldots, i_k) = (1, 2, \ldots, k)$ i $(i_1, \ldots, i_k) = (1, 2, \ldots, k, q)$. Załóżmy zatem, że chcemy to zrobić w pierwszej z sytuacji, czyli $(i_1, \ldots, i_k) = (1, 2, \ldots, k)$. Skoro $i_1 = 1, \ldots, i_k = k$, to:

$$
\begin{cases}
s_1 = 0, & v_1 = z_1, \\
s_2 = 0, & v_2 = z_1 + z_2, \\
\vdots & \\
s_m = 0, & v_m = z_1 + \ldots + z_m, \\
\vdots & \\
s_r = 0, & v_r = k, \ldots, n.
\end{cases}
$$

Dodatkowo założyliśmy, że $s_{r+1} = n - v_r$, $v_{r+1} = n$, $t_{z_1+\ldots+z_{r+1}} = \infty$, zatem:

$$
\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}) = \sum_{s_1=0}^{i_1-1} \sum_{v_1=i_{z_1}}^{i_{z_1+1}-1} \sum_{s_2=0}^{i_{z_1+1}-1-v_1} \sum_{v_2=i_{z_1+z_2}}^{i_{z_1+z_2+1}-1} \cdot \ldots \cdot \sum_{s_r=0}^{i_{k-z_r+1}-1-v_{r-1}} \times
$$

$$
\times \sum_{v_r=i_k}^{n} \prod_{j=1}^{r+1} \frac{n!}{s_j!(v_j - v_{j-1} - s_j)!} [F(t^-_{i_{z_1+\ldots+z_j}}) - F(t_{i_{z_1+\ldots+z_{j-1}}})]^{s_j} [p(t_{i_{z_1+\ldots+z_j}})]^{v_j - v_{j-1} - s_j} =
$$

$$
= \sum_{v_r=k}^{n} \underbrace{\prod_{j=1}^{r+1} \frac{n!}{s_j!(v_j - v_{j-1} - s_j)!}}_{(1*)} \underbrace{\prod_{j=1}^{r+1} [F(t^-_{z_1+\ldots+z_j}) - F(t_{z_1+\ldots+z_{j-1}})]^{s_j}}_{(2*)} \underbrace{\prod_{j=1}^{r+1} [p(t_{z_1+\ldots+z_j})]^{v_j - v_{j-1} - s_j}}_{(3*)} = (*).
$$

Rozpisując poszczególne fragmenty powyższego wyrażenia otrzymujemy:

$$
(1*) = \prod_{j=1}^{r+1} \frac{n!}{s_j!(v_j - v_{j-1} - s_j)!} = \frac{n!}{z_1! \cdot \ldots \cdot z_{r-1}!(v_r - k + z_r)!(n - v_r)!}.
$$

W wyrażeniu $(2*)$ jedynie czynnik odpowiadający $j = r + 1$ jest różny od 1, zatem:

$$
(2*) = \prod_{j=1}^{r+1} [F(t^-_{z_1+\ldots+z_j}) - F(t_{z_1+\ldots+z_{j-1}})]^{s_j} = [1 - F(t_k)]^{n-v_r}.
$$

Ostatni fragment wyrażenia wynosi:

$$
(3*) = \prod_{j=1}^{r+1} [p(t_{z_1+\ldots+z_j})]^{v_j - v_{j-1} - s_j} = [p(t_{z_1})]^{z_1} \cdot \ldots \cdot [p(t_{k-z_r})]^{z_{r-1}} [p(t_k)]^{v_r - k + z_r} = \prod_{i=1}^{k-z_r} p(t_i) [p(t_k)]^{v_r - k + z_r}.
$$

Zatem:

$$
(*) = \sum_{v_r=k}^{n} \frac{n!}{\prod_{j=1}^{r-1} z_j!(v_r - k + z_r)!(n - v_r)!} \prod_{i=1}^{k-z_r} p(t_i) [p(t_k)]^{v_r - k + z_r} [1 - F(t_k)]^{n - v_r} = (**).
$$

Aby zobaczyć, że powyższy wzór jest wzorem z twierdzenia 2.1, wystarczy przenumerować wyrazy sumy, podstawiając $s = v_r - k$. Wtedy:

$$
(**) = \sum_{s=0}^{n-k} \frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r + s)!(n - k - s)!} \prod_{i=1}^{k-z_r} p(t_i) [p(t_k)]^{z_r + s} [1 - F(t_k)]^{n-k-s} =
$$

$$
= \sum_{s=0}^{n-k} \frac{n!}{\prod_{j=1}^{r-1} z_j!(z_r + s)!(n - k - s)!} \prod_{i=1}^{k} p(t_i) [p(t_k)]^{s} [1 - F(t_k)]^{n-k-s}.
$$

Postępując analogicznie moglibyśmy otrzymać wzór z twierdzenia 2.4.

2.4.2 Przypadek zmiennych niezależnych o dowolnych rozkładach

Na końcu zajmiemy się przypadkiem najbardziej ogólnym, czyli zmiennymi niezależnymi, ale o dowolnych rozkładach dyskretnych. Poniższe twierdzenie jest zatem analogiem twierdzenia 2.6, z osłabionymi założeniami: zmienne losowe na podstawie których tworzone są statystyki porządkowe mogą mieć dowolne rozkłady.

TWIERDZENIE 2.7.

Niech $X_i \sim F_i, p_i$ - niezależne zmienne losowe oraz $t_{i_1}, \ldots, t_{i_k}$ ma r związanych bloków. Przyjmijmy dodatkowo, że $v_0 = 0$, $v_{r+1} = n$, $s_{r+1} = n - v_r$, $t_{i_{z_1+\ldots+z_0}} = -\infty$, $t_{i_{z_1+\ldots+z_{r+1}}} = \infty$.

Wtedy:

$$\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}) = \sum_{\substack{(j_1,\ldots,j_n) \\ \text{- permutacje zb.} \\ \{1,\ldots,n\}}} \sum_{s_1=0}^{i_1-1} \sum_{v_1=i_{z_1}}^{i_{z_1+1}-1} \sum_{s_2=0}^{i_{z_1+1}-1-v_1} \sum_{v_2=i_{z_1+z_2}}^{i_{z_1+z_2+1}-1} \cdot \ldots \cdot \sum_{s_r=0}^{i_{k-z_r+1}-1-v_{r-1}} \times$$

$$\times \sum_{v_r=i_k}^{n} \prod_{b=1}^{r+1} \frac{1}{s_b!(v_b - v_{b-1} - s_b)!} \prod_{a=v_{b-1}+1}^{v_{b-1}+s_b} [F_{j_a}(t^-_{i_{z_1+\ldots+z_b}}) - F_{j_a}(t_{i_{z_1+\ldots+z_{b-1}}})] \prod_{a=v_{b-1}+s_b+1}^{v_b} p_{j_a}(t_{i_{z_1+\ldots+z_b}}).$$

Dowód:

Dowód jest podobny do dowodów twierdzeń 2.3 i 2.6. Tak jak w twierdzeniu 2.6 szukamy prawdopodobieństwa $\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k})$ i zdefiniujemy w tym celu zdarzenia w podobny sposób jak w dowodzie twierdzenia 2.3.

$$P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)} : \begin{cases} X_{j_1}, \ldots, X_{j_{s_1}} < t_{i_{z_1}}, & \Rightarrow s_1 \text{ obserwacji jest } < t_{i_{z_1}}, \\ X_{j_{s_1+1}}, \ldots, X_{j_{v_1}} = t_{i_{z_1}}, & \Rightarrow v_1 - s_1 \text{ obserwacji jest } = t_{i_{z_1}}, \\ \vdots & \\ X_{j_{v_{m-1}+1}}, \ldots, X_{j_{v_{m-1}+s_m}} \in & \\ \qquad \in (t_{i_{z_1+\ldots+z_{m-1}}}, t_{i_{z_1+\ldots+z_m}}), & \Rightarrow s_m \text{ obserwacji } \in (t_{i_{z_1+\ldots+z_{m-1}}}, t_{i_{z_1+\ldots+z_m}}), \\ X_{j_{v_{m-1}+s_m+1}}, \ldots, X_{j_{v_m}} = t_{i_{z_1+\ldots+z_m}}, & \Rightarrow v_m - v_{m-1} - s_m \text{ obserwacji jest } = t_{i_{z_1+\ldots+z_m}}, \\ \vdots & \\ X_{j_{v_{r-1}+1}}, \ldots, X_{j_{v_{r-1}+s_r}} \in (t_{i_{k-z_r}}, t_{i_k}), & \Rightarrow s_r \text{ obserwacji jest } \in (t_{i_{k-z_r}}, t_{i_k}), \\ X_{j_{v_{r-1}+s_r+1}}, \ldots, X_{j_{v_r}} = t_{i_k}, & \Rightarrow v_r - v_{r-1} - s_r \text{ obserwacji jest } = t_{i_k}, \\ X_{j_{v_r+1}}, \ldots, X_{j_n} > t_{i_k}, & \Rightarrow n - v_r \text{ obserwacji jest } > t_{i_k}, \end{cases}$$

dla $s_1, \ldots, s_r$ oraz $v_1, \ldots, v_r$ zdefiniowanych tak samo jak w dowodzie tw. 2.6 oraz $(j_1, \ldots, j_n)$ - dowolnej permutacji zbioru $\{1, \ldots, n\}$.

Wtedy:

$$\mathbb{P}(X_{(i_1)} = t_{i_1}, \ldots, X_{(i_k)} = t_{i_k}) = \mathbb{P}(\bigcup_{(j_1,\ldots,j_n)} \bigcup_{(\mathbf{s},\mathbf{v})} P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)}) = (*).$$

Dla ustalonej permutacji $(j_1, \ldots, j_n)$ i różnych s_i, v_i zdarzenia $P^{(s_i,v_i)}_{(j_1,\ldots,j_n)}$ są parami rozłączne, zatem:

$$(*) = \sum_{(\mathbf{s},\mathbf{v})} \mathbb{P}(\bigcup_{(j_1,\ldots,j_n)} P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)}) = (**).$$

Jednak dla różnych permutacji $(j_1, \ldots, j_n)$ zdarzenia $P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)}$ nie są rozłączne, ale powtarzają się. Zatem wystarczy zliczyć ile razy każde ze zdarzeń powtarza się. Zliczamy je tak samo jak w dowodzie twierdzenia 2.3 i otrzymujemy:

$$s_1!(v_1 - s_1)! \cdot \ldots \cdot s_m!(v_m - v_{m-1} - s_m)! \cdot \ldots \cdot s_r!(v_r - v_{r-1} - s_r)!(n - v_r)!.$$

Zatem:

$$(**) = \sum_{(\mathbf{s},\mathbf{v})} \sum_{(j_1,\ldots,j_n)} \frac{1}{s_1!(v_1 - s_1)! \cdot \ldots \cdot s_r!(v_r - v_{r-1} - s_r)!(n - v_r)!} \mathbb{P}(P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)}) = (\clubsuit).$$

Trzeba już tylko znaleźć $\mathbb{P}(P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)})$ i robimy to tak jak w dowodach twierdzeń 2.3 i 2.5, korzystając z tego, że X_i są niezależne.

$$\begin{aligned}\mathbb{P}(P^{(\mathbf{s},\mathbf{v})}_{(j_1,\ldots,j_n)}) = &\prod_{a=1}^{s_1}[F_{j_a}(t^-_{i_{z_1}})] \prod_{a=s_1+1}^{v_1}[p_{j_a}(t_{i_{z_1}})]\cdot\ldots\cdot \prod_{a=v_{m-1}+1}^{v_{m-1}+s_m}[F_{j_a}(t^-_{i_{z_1+\ldots+z_m}}) - F_{j_a}(t_{i_{z_1+\ldots+z_{m-1}}})]\times\\ &\times \prod_{a=v_{m-1}+s_m+1}^{v_m}[p_{j_a}(t_{i_{z_1+\ldots+z_m}})]\cdot\ldots\cdot \prod_{a=v_{r-1}+1}^{v_{r-1}+s_r}[F_{j_a}(t^-_{i_k}) - F_{j_a}(t_{i_{k-z_r}})]\times\\ &\times \prod_{a=v_{r-1}+s_r+1}^{v_r}[p_{j_a}(t_{i_k})] \prod_{a=v_r+1}^{n}[1 - F_{j_a}(t_{i_k})] = (***).\end{aligned}$$

Jeżeli dodatkowo przyjmiemy, że $v_0 = 0$, $v_{r+1} = n$, $s_{r+1} = n - v_r$, $t_{i_{z_1+\ldots+z_0}} = -\infty$, $t_{i_{z_1+\ldots+z_{r+1}}} = \infty$, to wtedy

$$(***) = \prod_{b=1}^{r+1}\prod_{a=v_{b-1}+1}^{v_{b-1}+s_b}[F_{j_a}(t^-_{i_{z_1+\ldots+z_b}}) - F_{j_a}(t_{i_{z_1+\ldots+z_{b-1}}})] \prod_{a=v_{b-1}+s_b+1}^{v_b} p_{j_a}(t_{i_{z_1+\ldots+z_b}}).$$

Zatem ostatecznie:

$$\begin{aligned}(\clubsuit) &= \sum_{(\mathbf{s},\mathbf{v})}\sum_{(j_1,\ldots,j_n)}\prod_{b=1}^{r+1}\frac{1}{s_b!(v_b - v_{b-1} - s_b)!}\prod_{a=v_{b-1}+1}^{v_{b-1}+s_b}[F_{j_a}(t^-_{i_{z_1+\ldots+z_b}}) - F_{j_a}(t_{i_{z_1+\ldots+z_{b-1}}})] \prod_{a=v_{b-1}+s_b+1}^{v_b} p_{j_a}(t_{i_{z_1+\ldots+z_b}}) =\\ &= \sum_{(j_1,\ldots,j_n)}\sum_{s_1=0}^{i_1-1}\sum_{v_1=i_{z_1}}^{i_{z_1+1}-1}\sum_{s_2=0}^{i_{z_1+1}-1-v_1}\sum_{v_2=i_{z_1+z_2}}^{i_{z_1+z_2+1}-1}\cdot\ldots\cdot\sum_{s_r=0}^{i_{k-z_r+1}-1-v_{r-1}}\sum_{v_r=i_k}^{n}\times\\ &\times\prod_{b=1}^{r+1}\frac{1}{s_b!(v_b - v_{b-1} - s_b)!}\prod_{a=v_{b-1}+1}^{v_{b-1}+s_b}[F_{j_a}(t^-_{i_{z_1+\ldots+z_b}}) - F_{j_a}(t_{i_{z_1+\ldots+z_{b-1}}})] \prod_{a=v_{b-1}+s_b+1}^{v_b} p_{j_a}(t_{i_{z_1+\ldots+z_b}}).\end{aligned}$$

□

3 Rozkłady warunkowe

3.1 Prawdopodobieństwo warunkowe $\mathbb{P}(X_{(q)} = t_q | X_{(1)} = t_1, \ldots, X_{(k)} = t_k)$

W tym rozdziale naszym celem jest znalezienie prawdopodobieństwa warunkowego. Zakładamy, że znamy wartości pierwszych k statystyk porządkowych i na tej podstawie szukamy prawdopodobieństwa, że dla $q > k$, q-ta statystyka porządkowa przyjmie daną wartość.

3.1.1 Przypadek zmiennych niezależnych o jednakowym rozkładzie

Tak jak w poprzednich rozdziałach zakładamy, że zmienne losowe X_i pochodzą z rozkładów dyskretnych. Oznaczymy dla $1 \leqslant k \leqslant n$: $\mathbb{X}_{(k)} = (X_{(1)}, \ldots, X_{(k)})$ oraz $\mathbb{t}_k = (t_1, \ldots, t_k)$ takie, że:

$$t_z = \ldots = t_{z_1} < t_{z_1+1} = \ldots = t_{z_1+z_2} < \ldots < t_{z_1,\ldots,z_{r-1}+1} = \cdots = t_{z_1+\ldots+z_r} = t_k.$$

Naszym celem będzie znalezienie rozkładu warunkowego:

$$\mathbb{P}(X_{(q)} = t_q | X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \mathbb{P}(X_{(q)} = t_q | \mathbb{X}_{(k)} = \mathbb{t}_k).$$

Tak jak poprzednio, najpierw znajdziemy ten rozkład przy założeniu, że $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$, a następnie osłabimy to założenie zakładając jedynie niezależność obserwacji. Postać szukanego rozkładu warunkowego w przypadku zmiennych niezależnych o jednakowym rozkładzie wyniknie z trzech poniższych lematów.

LEMAT 3.1.
Załóżmy, że $X_1,\ldots,X_n \overset{\text{iid}}{\sim} F,p$ *oraz ciąg* $t_1 \leqslant \ldots \leqslant t_k$ *ma* r *związanych bloków. Wtedy:*

$$\begin{aligned}
&\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k) = \\
&= \sum_{s=0}^{n-k} \frac{n!}{(k-z_r)!(z_r+s)!(n-k-s)!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+s}[1-F(t_k)]^{n-k-s}.
\end{aligned} \tag{3.1}$$

Dowód:
Szukamy prawdopodobieństwa tego, że dokładnie $k-z_r$ obserwacji będzie mniejszych niż t_k, co najmniej z_r będzie równych t_k, a reszta będzie większa od t_k. Szukamy zatem prawdopodobieństwa, że zajdzie zdarzenie:

$$Z : \begin{cases} k-z_r \text{ obserwacji jest mniejszych od } t_k, \\ \text{co najmniej } z_r \text{ obserwacji jest równych } t_k, \\ \text{pozostałe obserwacje są większe lub równe } t_k. \end{cases}$$

Zdarzenie to można zilustrować przy pomocy rysunku:

$$\underbrace{X_{(1)}\cdots X_{(k-z_r)}}_{<t_k}\ \underbrace{X_{(k-z_r+1)}\cdots X_{(k)}\cdots X_{(k+s)}}_{=t_k}\ \underbrace{X_{(k+s+1)}\cdots X_{(n)}}_{>t_k}$$

Zdefiniujmy zatem dla $s=0,\ldots,n-k$ zdarzenia:

$$A_s : \begin{cases} k-z_r & \text{obserwacji jest mniejszych od } t_k & (\text{ ozn. } B_{1s}), \\ z_r+s & \text{obserwacji jest równych } t_k & (\text{ ozn. } B_{2s}), \\ n-k-s & \text{obserwacji jest większych od } t_k & (\text{ ozn. } B_{3s}). \end{cases}$$

Wtedy

$$\mathbb{P}(Z) = \mathbb{P}(\bigcup_{s=0}^{n-k} A_s) = (*),$$

a ponieważ zdarzenia A_s są parami rozłączne, to

$$(*) = \sum_{s=0}^{n-k} \mathbb{P}(A_s).$$

Szukamy zatem prawdopodobieństwa $\mathbb{P}(A_s)$.

$$\begin{aligned}
\mathbb{P}(A_s) &= \mathbb{P}(B_{1s}, B_{2s}, B_{3s}) = \\
&= \mathbb{P}(B_{1s})\mathbb{P}(B_{2s}|B_{1s})\mathbb{P}(B_{3s}|B_{1s},B_{2s}) = \\
&= \binom{n}{k-z_r}[F(t_k^-)]^{k-z_r}\binom{n-k+z_r}{z_r+s}[p(t_k)]^{z_r+s}\binom{n-k+z_r-z_r-s}{n-k-s}[1-F(t_k)]^{n-k-s} = \\
&= \frac{n!}{\cancel{(n-k+z_r)!}(k-z_r)!}\frac{\cancel{(n-k+z_r)!}}{(n-k-s)!(z_r+s)!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+s}[1-F(t_k)]^{n-k-s} = \\
&= \frac{n!}{(k-z_r)!(n-k-s)!(z_r+s)!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+s}[1-F(t_k)]^{n-k-s}.
\end{aligned}$$

Ostatecznie otrzymujemy:

$$\begin{aligned}
&\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k) = \\
&= \sum_{s=0}^{n-k} \frac{n!}{(k-z_r)!(n-k-s)!(z_r+s)!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+s}[1-F(t_k)]^{n-k-s}.
\end{aligned}$$

□

LEMAT 3.2.

Tak jak w lemacie 3.1 załóżmy, że $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$ *oraz ciąg* $t_1 \leqslant \ldots \leqslant t_k$ *ma* r *związanych bloków. Dla* $q > k$ *zachodzi:*

$$\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k, X_{(q)} = t_q) =$$
$$= \begin{cases} \sum\limits_{s=0}^{n-q} \frac{n!}{(k-z_r)!(z_r+q-k+s)!(n-q-s)!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+q-k+s}[1-F(t_k)]^{n-q-s}, & gdy\ t_k = t_q, \\ \sum\limits_{u=0}^{q-k-1} \sum\limits_{h=0}^{q-k-1-u} \sum\limits_{v=0}^{n-q} \frac{n!}{(k-z_r)!(z_r+u)!h!v!(n-k-u-v-h)!} \times & \\ \quad \times [F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+u}[F(t_q^-) - F(t_k)]^h [p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v, & gdy\ t_k < t_q. \end{cases} \tag{3.2}$$

Dowód:

Najpierw zajmiemy się przypadkiem, gdy $t_k = t_q$. Jest do dokładnie taka sama sytuacja jak w lemacie 3.1, trzeba tylko uwzględnić we wzorze dodatkowe $q-k$ obserwacji, które są również równe t_k. Pokazuje to poniższy rysunek.

$$\underbrace{X_{(1)} \cdots X_{(k-z_r)}}_{<t_k}\ \underbrace{X_{(k-z_r+1)} \cdots X_{(k)} \cdots X_{(q)} \cdots X_{(q+s)}}_{=t_k}\ \underbrace{X_{(k+s+1)} \cdots X_{(n)}}_{>t_k}$$

Zatem korzystając z lematu 3.1, dla $t_k = t_q$ zachodzi:

$$\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k, X_{(q)} = t_q) =$$
$$= \sum_{s=0}^{n-q} \frac{n!}{(k-z_r)!(z_r+q-k+s)!(n-q-s)!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+q-k+s}[1-F(t_k)]^{n-q-s}. \tag{3.3}$$

Następnie rozważymy przypadek $t_k < t_q$. Sytuację najlepiej zilustrować rysunkiem.

$$\overbrace{\underbrace{X_{(1)} \cdots X_{(k-z_r)}}_{<t_k}}^{k-z_r}\ \overbrace{\underbrace{X_{(k-z_r+1)} \cdots X_{(k+u)}}_{=t_k}}^{z_r+u}\ \overbrace{\underbrace{X_{(k+u+1)} \cdots X_{(k+u+h)}}_{\in(t_k,t_q)}}^{h}\ \overbrace{\underbrace{X_{(k+u+h+1)} \cdots X_{(n-v)}}_{=t_q}}^{n-k-u-h-v}\ \overbrace{\underbrace{X_{(n-v+1)} \cdots X_{(n)}}_{>t_q}}^{v}$$

Zdefiniujmy zatem zdarzenia:

$$A_{uhv}: \begin{cases} k - z_r & \text{obserwacji jest mniejszych od } t_k \quad (\text{ ozn. } B_{1(uhv)}), \\ z_r + u & \text{obserwacji jest równych } t_k \quad (\text{ ozn. } B_{2(uhv)}), \\ h & \text{obserwacji leży pomiędzy } t_k \text{ i } t_q \quad (\text{ ozn. } B_{3(uhv)}), \\ n-k-v-u-h & \text{obserwacji jest równych } t_q \quad (\text{ ozn. } B_{4(uhv)}), \\ v & \text{obserwacji jest większych od } t_q \quad (\text{ ozn. } B_{5(uhv)}), \end{cases}$$

dla:

$v = 0, \ldots, n-q,$

$u = o, \ldots, q-k-1,$

$h = 0, \ldots, q-k-1-u.$

Wtedy $\forall u, v, h$ będzie $k - z_r$ obserwacji mniejszych niż t_k, co najmniej $z_r - 0$ obserwacji równych t_k oraz co najmniej $n - k - (n-q) - u - (q-k-1-u) = 1$ obserwacja równa t_q.
Dla tak zdefiniowanych zdarzeń A_{uhv} można zapisać:

$$\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k, X_q = t_q) = \mathbb{P}(\bigcup_{v=0}^{n-q}\bigcup_{u=0}^{q-k-1}\bigcup_{h=0}^{q-k-1-u} A_{uhv}) = (*).$$

Wszystkie zdarzenia A_{uhv} są parami rozłączne, zatem:

$$(*) = \sum_{v=0}^{n-q}\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-1-u} \mathbb{P}(A_{uhv}).$$

Wystarczy teraz znaleźć $\mathbb{P}(A_{uhv})$.

$$\begin{aligned}
\mathbb{P}(A_{uhv}) &= \mathbb{P}(B_1, B_2, B_3, B_4, B_5) = \\
&= \mathbb{P}(B_1)\mathbb{P}(B_2|B_1)\mathbb{P}(B_3|B_1, B_2)\mathbb{P}(B_4|B_1, B_2, B_3)\mathbb{P}(B_5|B_1, B_2, B_3, B_4) = \\
&= \binom{n}{k-z_r}[F(t_k^-)]^{k-z_r}\binom{n-k+z_r}{z_r+u}[p(t_k)]^{z_r+u}\binom{n-k-u}{h}[F(t_q^-) - F(t_k)]^h \times \\
&\quad \times \binom{n-k-u-h}{n-k-u-h-v}[p(t_q)]^{n-k-u-h-v}\binom{v}{v}[F(t_q)]^v = \\
&= \frac{n!}{(k-z_r)!\cancel{(n-k+z_r)!}}\frac{\cancel{(n-k+z_r)!}}{(z_r+u)!\cancel{(n-k-u)!}}\frac{\cancel{(n-k-u)!}}{h!\cancel{(n-k-u-h)!}}\frac{\cancel{(n-k-u-h)!}}{(n-k-u-h-v)!v!} \times \\
&\quad \times [F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+u}[F(t_q^-) - F(t_k)]^h[p(t_q)]^{n-k-u-h-v}[F(t_q)]^v = \\
&= \frac{n!}{(k-z_r)!(z_r+u)!h!(n-k-u-h-v)!v!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+u}[F(t_q^-) - F(t_k)]^h \times \\
&\quad \times [p(t_q)]^{n-k-u-h-v}[F(t_q)]^v.
\end{aligned}$$

Zatem ostatecznie:

$$\begin{aligned}
&\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k, X_q = t_q) = \\
&= \sum_{v=0}^{n-q}\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-1-u} \frac{n!}{(k-z_r)!(z_r+u)!h!(n-k-u-h-v)!v!}[F(t_k^-)]^{k-z_r}[p(t_k)]^{z_r+u}[F(t_q^-) - F(t_k)]^h \times \\
&\quad \times [p(t_q)]^{n-k-u-h-v}[F(t_q)]^v.
\end{aligned}$$

□

Ostatnim faktem potrzebnym do znalezienia szukanej postaci rozkładu jest poniższy lemat.

LEMAT 3.3.

Tak jak w lematach 3.1 i 3.2 załóżmy, że $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, p$ oraz ciąg $t_1 \leqslant \ldots \leqslant t_k$ ma r związanych bloków. Dla $q > k$ zachodzi:

$$\mathbb{P}(X_{(q)} = t_q | \mathbb{X}_{(k)} = \mathbb{t}_k) = \mathbb{P}(X_{(q)} = t_q | X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k).$$

Dowód:

Należy rozważyć 2 przypadki: $t_k < t_q$ oraz $t_k = t_q$.

Zajmiemy się najpierw pierwszym z nich. Zapiszmy lewą stronę nierówności.

$$\begin{aligned}
\mathrm{L} &= \mathbb{P}(X_{(q)} = t_q | X_{(1)} = t_1, \ldots, X_{(k)} = t_k) = \\
&= \frac{\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k, X_{(q)} = t_q)}{\mathbb{P}(X_{(1)} = t_1, \ldots, X_{(k)} = t_k)} = (*).
\end{aligned} \tag{3.4}$$

Korzystając z twierdzeń 2.1 i 2.4 możemy zapisać postać licznika i mianownika:

$$(*) = \frac{\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-1-u}\sum_{v=0}^{n-q}\frac{\cancel{n!}\cancel{[\prod_{j=1}^{k}p(t_j)]}[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v}{\cancel{\prod_{j=1}^{r-1}z_j!}(z_r+u)!h!(n-k-u-h-v)!v!}}{\sum_{s=0}^{n-k}\frac{\cancel{n!}\cancel{[\prod_{i=1}^{k}p(t_i)]}[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(n-k-s)!(z_r+s)!\cancel{\prod_{j=1}^{r-1}z_j!}}} =$$

$$= \frac{\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-1-u}\sum_{v=0}^{n-q}\frac{[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v}{(z_r+u)!h!(n-k-u-h-v)!v!}}{\sum_{s=0}^{n-k}\frac{[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(n-k-s)!(z_r+s)!}}.$$

Korzystając z lematów 3.1 i 3.2 możemy rozpisać prawą stronę.

$$\mathrm{P} = \frac{\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = \cdots = X_{(k)} = t_k, X_{(q)} = t_q)}{\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = \cdots = X_{(k)} = t_k)} =$$

$$= \frac{\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-1-u}\sum_{v=0}^{n-q}\frac{\cancel{n!}\cancel{[F(t_k^-)]^{k-z_r}}[p(t_k)]^{\cancel{z_r}+u}[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v}{\cancel{(k-z_r)!}(z_r+u)!h!v!(n-k-u-v-h)!}}{\sum_{s=0}^{n-k}\frac{\cancel{n!}\cancel{[F(t_k^-)]^{k-z_r}}[p(t_k)]^{\cancel{z_r}+s}[1-F(t_q)]^{n-k-s}}{\cancel{(k-z_r)!}(z_r+s)!(n-k-s)!}}$$

$$= \frac{\sum_{u=0}^{q-k-1}\sum_{h=0}^{q-k-1-u}\sum_{v=0}^{n-q}\frac{[p(t_k)]^u[F(t_q^-)-F(t_k)]^h[p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v}{(z_r+u)!h!v!(n-k-u-v-h)!}}{\sum_{s=0}^{n-k}\frac{[p(t_k)]^s[1-F(t_q)]^{n-k-s}}{(z_r+s)!(n-k-s)!}}.$$

Zatem L=P.

Analogicznie postępujemy, gdy $t_k = t_q$. Rozpisujemy najpierw lewą stronę równości:

$$\mathrm{L} = \frac{\sum_{v=0}^{n-q}\frac{\cancel{n!}}{\cancel{\prod_{j=1}^{r-1}z_j!}(z_r+q-k+v)!(n-k-v)!}\cancel{[\prod_{j=1}^{k}p(t_j)]}[p(t_k)]^v[1-F(t_k)]^{n-k-v}}{\sum_{s=0}^{n-k}\frac{\cancel{n!}}{(n-k-s)!(z_r+s)!\cancel{\prod_{j=1}^{r-1}z_j!}}\cancel{[\prod_{i=1}^{k}p(t_i)]}[p(t_k)]^s[1-F(t_k)]^{n-k-s}} =$$

$$= \frac{\sum_{v=0}^{n-q}\frac{[p(t_k)]^v[1-F(t_k)]^{n-k-v}}{(z_r+q-k+v)!(n-k-v)!}}{\sum_{s=0}^{n-k}\frac{[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(n-k-s)!(z_r+s)!}}.$$

A następnie prawą stronę:

$$\mathrm{P} = \frac{\sum_{s=0}^{n-q}\frac{\cancel{n!}}{\cancel{(k-z_r)!}(z_r+q-k+s)!(n-q-s)!}\cancel{[F(t_k^-)]^{k-z_r}}[p(t_k)]^{\cancel{z_r}+q-k+s}[1-F(t_k)]^{n-q-s}}{\sum_{s=0}^{n-k}\frac{\cancel{n!}}{\cancel{(k-z_r)!}(z_r+s)!(n-k-s)!}\cancel{[F(t_k^-)]^{k-z_r}}[p(t_k)]^{\cancel{z_r}+s}[1-F(t_k)]^{n-k-s}} =$$

$$= \frac{\sum_{s=0}^{n-q}\frac{[p(t_k)]^{q-k+s}[1-F(t_k)]^{n-q-s}}{(z_r+q-k+s)!(n-q-s)!}}{\sum_{s=0}^{n-k}\frac{[p(t_k)]^s[1-F(t_k)]^{n-k-s}}{(z_r+s)!(n-k-s)!}}.$$

Żeby zobaczyć, że L=P, przenumerujemy indeksy w sumie z licznika powyższego wyrażenia.

$$v := n-q-s \;\Rightarrow\; s = n-q-s,$$
$$s = 0 \;\Rightarrow\; v = n-q,$$
$$s = n-q \;\Rightarrow\; v = 0.$$

Wtedy:

$$P = \frac{\sum_{v=0}^{n-q} \frac{[p(t_k)]^v [1-F(t_k)]^{n-k-v}}{(z_r+q-k+v)!(n-k-v)!}}{\sum_{s=0}^{n-k} \frac{[p(t_k)]^s [1-F(t_k)]^{n-k-s}}{(n-k-s)!(z_r+s)!}} = L.$$

Zatem $L = P$.

□

WNIOSEK 3.1.
Wiadomo, że statystyki porządkowe utworzone na podstawie zmiennych losowych z rozkładu dyskretnego, który ma co najmniej 3 punkty w nośniku nie posiadają własności Markowa, czyli dla $2 \leqslant k < q \leqslant n$ nie zachodzi:

$$\mathbb{P}(X_{(q)} = t_q | X_{(k)} = t_k, \dots, X_{(1)} = t_1) \neq \mathbb{P}(X_{(q)} = t_q | X_{(k)} = t_k).$$

Jest to pokazane na przykład w [2].
Lemat 3.3 pokazuje jednak, że jeśli do warunku po prawej stronie równości dodamy informację o tym jak długi jest ostatni związany blok obserwacji, to równość będzie prawdziwa, więc

$$\mathbb{P}(X_{(q)} = t_q | X_{(k)} = t_k, \dots, X_{(1)} = t_1) = \mathbb{P}(X_{(q)} = t_q | X_{(k)} = t_k, \text{ ostatni związany blok obserwacji ma długość } z_r) =$$
$$= \mathbb{P}(X_{(q)} = t_q | X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \dots = X_{(k)} = t_k).$$

Po udowodnieniu lematów: 3.1, 3.2 oraz 3.3 możemy znaleźć już szukane prawdopodobieństwo warunkowe.

TWIERDZENIE 3.1.
Załóżmy, że $X_1, \dots, X_n \overset{iid}{\sim} F, p$ *oraz ciąg* $t_1 \leqslant \dots \leqslant t_k$ *ma* r *związanych bloków. Dla* $1 \leqslant k < q \leqslant n$, $p(t_j) > 0$ $\forall j = 1, \dots, k$, *zachodzi:*

$$\mathbb{P}(X_{(q)} = t_q | \mathbb{X}_{(k)} = \mathbb{t}_k) = \begin{cases} \frac{A}{B} & \text{dla } t_k = t_q, \\ \frac{C}{D} & \text{dla } t_k < t_q, \end{cases}$$

gdzie:

$$A = \sum_{v=0}^{n-q} \frac{1}{(z_r + n - k - v)! v!} \left(\frac{1 - F(t_k)}{p(t_k)} \right)^v,$$
$$B = \sum_{v=0}^{n-k} \frac{1}{(z_r + n - k - v)! v!} \left(\frac{1 - F(t_k)}{p(t_k)} \right)^v,$$
$$C = \sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-1-u} \sum_{v=0}^{n-q}{}' \frac{1}{(z_r + u)! h! v! (n - k - u - v - h)!} [p(t_k)]^{z_r+u} [F(t_q^-) - F(t_k)]^h [p(t_q)]^{n-k-u-h-v} [1 - F(t_q)]^v,$$
$$D = \sum_{s=0}^{n-k} \frac{1}{(z_r + s)!(n - k - s)!} [p(t_k)]^s [1 - F(t_k)]^{n-k-s}.$$

Dowód:
Korzystając z lematu 3.3 można zapisać:

$$\mathbb{P}(X_{(q)} = t_q | \mathbb{X}_{(k)} = \mathbb{t}_k) = \mathbb{P}(X_{(q)} = t_q | X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \dots = X_{(k)} = t_k) =$$
$$= \frac{\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \dots = X_{(k)} = t_k, X_{(q)} = t_q)}{\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \dots = X_{(k)} = t_k)} = (*). \quad (3.5)$$

Rozważmy najpierw przypadek $t_k < t_q$. Wtedy korzystając z lematów 3.1 i 3.2 oraz z powyższej równości możemy napisać:

$$(*) = \frac{\sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-1-u} \sum_{v=0}^{n-q} \cancel{\frac{n!}{(k-z_r)!}}\frac{1}{(z_r+u)!h!v!(n-k-u-v-h)!} \cancel{[F(t_k^-)]^{k-z_r}} [p(t_k)]^{\cancel{z_r}+u}[1-F(t_q)]^v}{\sum_{s=0}^{n-k} \cancel{\frac{n!}{(k-z_r)!}}\frac{1}{(z_r+s)!(n-k-s)!} \cancel{[F(t_k^-)]^{k-z_r}} [p(t_k)]^{\cancel{z_r}+s}[1-F(t_k)]^{n-k-s}} \times$$

$$\times \frac{[F(t_q^-)-F(t_k)]^h [p(t_q)]^{n-k-u-h-v}}{1} =$$

$$= \frac{\sum_{u=0}^{q-k-1} \sum_{h=0}^{q-k-1-u} \sum_{v=0}^{n-q} \frac{1}{(z_r+u)!h!v!(n-k-u-v-h)!} [p(t_k)]^u [F(t_q^-)-F(t_k)]^h [p(t_q)]^{n-k-u-h-v}[1-F(t_q)]^v}{\sum_{s=0}^{n-k} \frac{1}{(z_r+s)!(n-k-s)!}[p(t_k)]^s[1-F(t_k)]^{n-k-s}} = \frac{C}{D}.$$

Następnie spójrzmy na przypadek $t_k = t_q$. Ponownie korzystając z równości 3.5 zapisujemy:

$$(*) = \frac{\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k = \cdots = X_{(q)} = t_k)}{\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k)} = (**).$$

Tym razem zastosujemy lemat 3.1 zarówno do licznika, jak i do mianownika i otrzymamy:

$$(**) = \frac{\sum_{s=0}^{n-q} \cancel{\frac{n!}{(k-z_r)!}}\frac{1}{(z_r+q-k+s)!(n-q-s)!} \cancel{[F(t_k^-)]^{k-z_r}} [p(t_k)]^{\cancel{z_r}+q-k+s}[1-F(t_k)]^{n-q-s}}{\sum_{s=0}^{n-k} \cancel{\frac{n!}{(k-z_r)!}}\frac{1}{(z_r+s)!(n-k-s)!} \cancel{[F(t_k^-)]^{k-z_r}} [p(t_k)]^{\cancel{z_r}+s}[1-F(t_k)]^{n-k-s}} =$$

$$= \frac{\sum_{s=0}^{n-q} \frac{1}{(z_r+q-k+s)!(n-q-s)!}[p(t_k)]^{q-k+s}[1-F(t_k)]^{n-q-s}}{\sum_{s=0}^{n-k} \frac{1}{(z_r+s)!(n-k-s)!}[p(t_k)]^s[1-F(t_k)]^{n-k-s}} = (***).$$

Aby przedstawić powyższy wynik w takiej postaci jak w tezie twierdzenia, wystarczy przenumerować wyrazy w sumach.
W liczniku podstawiamy $v := n-q-s$:

$s \quad \to n-q-v,$

$z_r+q-k+s \quad \to z_r+\cancel{q}-k+n-\cancel{q}-v = z_r-k+n-v,$

$q-k+s \quad \to \cancel{q}-k+n-\cancel{q}-v = n-k-v,$

$s=0 \quad \to v=n-q,$

$s=n-q \quad \to v=0.$

A w mianowniku podstawiamy $v := n-k-s$:

$s \quad \to n-k-v,$

$z_r+s \quad \to z_r+n-k-v,$

$s=0 \quad \to v=n-k,$

$s=n-k \quad \to v=0.$

Ostatecznie otrzymujemy:

$$(***) = \frac{\sum_{v=0}^{n-q} \frac{1}{(z_r+n-k-v)!v!}[p(t_k)]^{\cancel{n-k}-v}[1-F(t_k)]^v}{\sum_{v=0}^{n-k} \frac{1}{(z_r+n-k-v)!v!}[p(t_k)]^{\cancel{n-k}-v}[1-F(t_k)]^v} = \frac{\sum_{v=0}^{n-q} \frac{1}{(z_r+n-k-v)!v!}\left(\frac{1-F(t_k)}{p(t_k)}\right)^v}{\sum_{v=0}^{n-k} \frac{1}{(z_r+n-k-v)!v!}\left(\frac{1-F(t_k)}{p(t_k)}\right)^v} = \frac{A}{B}.$$

$\square$

3.1.2 Przypadek zmiennych niezależnych o dowolnych rozkładach

Kolejną sytuacją jest rozważenie zmiennych niezależnych o dowolnych rozkładach. Okazuje się, że nie da się udowodnić faktów analogicznych do tych z poprzedniego podrozdziału. Można pokazać, że lematy 3.1 i 3.2 są prawdziwe przy założeniu, że $X_i \sim F_i, p_i$, jednak lemat 3.3 nie jest prawdziwy w tym przypadku, co pokażemy na przykładzie.
Aby pokazać kontrprzykład trzeba najpierw wykazać prawdziwość faktu analogicznego do lematu 3.1.

LEMAT 3.4.
Załóżmy, że $X_i \sim F_i, p_i$ dla $i = 1, \ldots, n$ są niezależnymi zmiennymi losowymi oraz ciąg $t_1 \leqslant \ldots \leqslant t_k$ ma r związanych bloków. Wtedy:

$$\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k) =$$
$$= \sum_{s=0}^{n-k} \sum_{::} \frac{n!}{(k-z_r)!(z_r+s)!(n-k-s)!} \prod_{i=1}^{k-z_r} [F_{j_i}(t_k^-)] \prod_{i=k-z_r+1}^{k+s} [p_{j_i}(t_k)] \prod_{i=k+s+1}^{n} [1 - F_{j_i}(t_k)]. \tag{3.6}$$

Dowód:
Dowód jest podobny do dowodów twierdzeń dotyczących zmiennych o dowolnych rozkładach oraz dowodu lematu 3.1. Zdefiniujmy zatem zdarzenia:

$$P^{(s)}_{(j_1,\ldots,j_n)} : \begin{cases} X_{j_1}, \ldots, X_{j_{k-z_r}} < t_k, \\ \vdots \\ X_{j_{k-z_r+1}}, \ldots, X_{j_{k+s}} = t_k, \\ X_{j_{k+s+1}} = \ldots = X_{j_n} > t_k. \end{cases}$$

gdzie $(j_1, \ldots, j_n)$ - dowolna permutacja $\{1, \ldots, n\}$

Wtedy:

$$\mathbb{P}(X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k) = \sum_{s=0}^{n-k} (\mathbb{P}(\bigcup_{(j_1,\ldots,j_n)} P^{(s)}_{(j_1,\ldots,j_n)})) = (*).$$

Suma $\bigcup_{(j_1,\ldots,j_n)} P^{(s)}_{(j_1,\ldots,j_n)}$ składa się z powtarzających się zdarzeń. Każde ze zdarzeń powtarza się w sumie $\bigcup_{(j_1,\ldots,j_n)} P^{(s)}_{(j_1,\ldots,j_n)}$ dokładnie $(k-z_r)!(z_r+s)!(n-k-s)!$ razy. Zatem:

$$(*) = \sum_{s=0}^{n-k} \sum_{(j_1,\ldots,j_n)} (\frac{1}{(k-z_r)!(z_r+s)!(n-k-s)!} \mathbb{P}(P^{(s)}_{(j_1,\ldots,j_n)})) = (**).$$

Pozostaje znalezienie $\mathbb{P}(P^{(s)}_{(j_1,\ldots,j_n)})$. Korzystając z niezależności obserwacji X_i otrzymujemy:

$$\mathbb{P}(P^{(s)}_{(j_1,\ldots,j_n)}) = \prod_{i=1}^{k-z_r} F_{j_i}(t_k^-) \prod_{i=k-z_r+1}^{k+s} p_{j_i}(t_k) \prod_{i=k+s+1}^{n} [1 - F_{j_i}(t_k)].$$

Zatem ostatecznie otrzymujemy tezę lematu:

$$(**) = \sum_{s=0}^{n-k} \sum_{(j_1,\ldots,j_n)} \prod_{i=1}^{k-z_r} F_{j_i}(t_k^-) \prod_{i=k-z_r+1}^{k+s} p_{j_i}(t_k) \prod_{i=k+s+1}^{n} [1 - F_{j_i}(t_k)].$$

□

Pokażemy teraz, że nie jest prawdziwy analog lematu 3.3 dla zmiennych o dowolnych rozkładach.

KONTRPRZYKŁAD 3.1.

Pokażemy, że dla niezależnych zmiennych $X_i \sim F_i, p_i$, w sytuacji gdy F_i, p_i mogą być różne, nie musi zachodzić:

$$\mathbb{P}(X_{(q)} = t_q | \mathbb{X}_{(k)} = \mathbf{t}_k) = \mathbb{P}(X_{(q)} = t_q | X_{(k-z_r)} < X_{(k-z_r+1)} = X_{(k-z_r+2)} = \cdots = X_{(k)} = t_k). \tag{3.7}$$

Niech $n = q = 3$, $k = 2$, $r = 2$, $z_1 = 1$, $z_2 = 1$, $t_k = t_q$ oraz $t_1 = 4$, $t_2 = 5$, $t_3 = t_2$. Rozważmy różne rozkłady geometryczne:

$$X_1 \sim geom(0.4) \Rightarrow p_1(x) = \frac{4}{10}\left(\frac{6}{10}\right)^x, F_1(x) = 1 - \left(\frac{6}{10}\right)^x$$
$$X_2 \sim geom(0.8) \Rightarrow p_2(x) = \frac{8}{10}\left(\frac{2}{10}\right)^x, F_2(x) = 1 - \left(\frac{2}{10}\right)^x$$
$$X_3 \sim geom(0.6) \Rightarrow p_3(x) = \frac{6}{10}\left(\frac{4}{10}\right)^x, F_3(x) = 1 - \left(\frac{4}{10}\right)^x$$

Wtedy wzór 3.7 przyjmuje postać:

$$\mathbb{P}(X_{(3)} = t_3 | X_{(1)} = t_1, X_{(2)} = t_2) = \mathbb{P}(X_{(3)} = t_3 | X_{(1)} < X_{(2)} = t_2).$$

Rozpiszemy osobno lewą i prawą stronę równania, korzystając ze wzorów z twierdzenia 2.3 i lematu 3.4 odpowiednio:

$$\begin{aligned}
\mathrm{L} &= \frac{\mathbb{P}(X_{(1)} = t_1, X_{(2)} = t_2, X_{(3)} = t_2)}{\mathbb{P}(X_{(1)} = t_1, X_{(2)} = t_2)} = \\
&= \frac{\sum\limits_{(j_1,j_2,j_3)} \frac{1}{2} p_{j_1}(t_1) p_{j_2}(t_2) p_{j_3}(t_2)}{\sum\limits_{(j_1,j_2,j_3)} \left(\frac{1}{2} p_{j_1}(t_1) p_{j_2}(t_2) p_{j_3}(t_2) + p_{j_1}(t_1) p_{j_2}(t_2)[1 - F_{j_3}(t_2)]\right)} = (L*) \\
\mathrm{P} &= \frac{\mathbb{P}(X_{(1)} < X_{(2)} = t_2, X_{(3)} = t_3)}{\mathbb{P}(X_{(1)} < X_{(2)} = t_2)} = \\
&= \frac{\sum\limits_{(j_1,j_2,j_3)} \frac{1}{2} F_{j_1}(t_2^-) p_{j_2}(t_2) p_{j_3}(t_2)}{\sum\limits_{(j_1,j_2,j_3)} \left(\frac{1}{2} F_{j_1}(t_2^-) p_{j_2}(t_2) p_{j_3}(t_2) + F_{j_1}(t_2^-) p_{j_2}(t_2)[1 - F_{j_3}(t_2)]\right)} = (P*)
\end{aligned}$$

Powyższe obliczenia zostały wykonane przy pomocy pakietu R i otrzymano wyniki:

$$(L*) \approx 0.3539$$
$$(P*) \approx 0.3183$$

Zatem widać, że dla powyższego przykładu równość nie zachodzi.

WNIOSEK 3.2.

Dla niezależnych zmiennych dyskretnych o dowolnych rozkładach nie zachodzi, ani własność Markowa, ani też, w przeciwieństwie do zmiennych o jednakowych rozkładach, warunek z lematu 3.3.

Literatura

[1] G. Gan, L.J. Bain, *Distribution of order statistics for discrete parents with applications to censored sampling*, Journal of Statistical Planning and Inference **44**, 37-46, (1995).

[2] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, *A First Course in Order Statistics*, John Wiley & Sons Inc. (1992).

[3] N. Balakrishnan, *Permanents, Order Statistics, Outliers and Robustness*, Revista Matematica Complutense **20**, 7-107, (2007).

[4] A. Dembińska, *Charakteryzacje rozkładów prawdopodobieństwa związane z własnościami regresyjnymi statystyk porządkowych i rekordowych*, rozprawa doktorska na Wydziale MiNI PW, (2002).

[5] T. Rychlik *Bias-Robustness of L-estimates of location against dependence*, Statistics **24**, 9-15, (1993).