Универзитет у Београду

Математички факултет

Тачне расподеле и моменти статистика поретка

Мастер рад

Ментор:

др Павле Младеновић

Студент:

Наташа Иветић

Београд, 2015.

Садржај

Увод

Статистике поретка имају велики значај у вероватноћи и статистици, актуарској и финансијској математици као и у многим другим областима. Статистикама поретка је посвећено доста пажње, посебно у последњих неколико година. На ову тему је објављен велики број научних радова, као и књига. Многи аутори су се бавили њиховим особинама, расподелама, моментима, асимптотском теоријом екстремних статистика поретка и њиховом применом у тестирању статистичких хипотеза и оцењивању параметара, затим границама и апроксимацијом момената статистика поретка, као и случајеве кад су оне равномерно распоређене и када нису. Као што видимо статистике поретка предтављају обимну тему за проучавање. У раду се разматрају проблеми који се односе на тачне расподеле статистика поретка и на проблеме одређивања њихових момената. А делом је захваћена и асимтотска теорија статистика поретка, где су дате оцене индекса репа расподеле које се дефинишу помоћу екстеремних статистике поретка.

1. Теорија расподела статистика поретка

1.1. Дефиниција статистика поретка и њихова својства

Низ који се добија када се кад се подаци из узорка уреде у неопадајућем поретку назива се варијациони низ, а елементи тог низа називају се статистике поретка. Дакле, нека је $(X_1, X_2, \ldots, X_n)$ случајан узорак и нека је $(x_1, x_2, \ldots, x_n)$ једна реализација тог узорка. Бројеве $x_1, x_2, \ldots, x_n$ поређамо по величини и добијемо низ $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. Случајне величине дефинишемо на следећи начин, за сваку реализацију $(x_1, x_2, \ldots, x_n)$ нека је

$$X_{(1)} = x_{(1)}, \ X_{(2)} = x_{(2)}, \ \ldots, \ X_{(n)} = x_{(n)}.$$

Случајне величине $X_{(1)}, X_{(2)}, \ \ldots \ , X_{(n)}$ зову се статистике поретка, а низ $X_{(1)} \leq X_{(2)} \leq \ \ldots \ \leq X_{(n)}$ зове се варијациони низ узорка $(X_1, X_2, \ldots, X_n)$. Дакле, статистике поретка $X_{(1)}, X_{(2)}, \ \ldots \ , X_{(n)}$ првог, другог, ... n-тог ранга су редом први, други, ... , n-ти елемент варијационог низа. Посебно се издвајају:

статистика поретка првог ранга $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$,

статистика поретка n-тог ранга $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

Помоћу њих се дефинише распон узорка $W = X_{(n)} - X_{(1)}$.

Статистике поретка се користе за дефинисање медијане узорка. Узорачка медијана је статистика

$$M_e = \begin{cases} X_{k+1}, & ako\ je\ n = 2k+1 \\ \frac{1}{2}(X_k + X_{k+1}), & ako\ je\ n = 2k \end{cases}.$$

Видимо да је за непарно n медијана такође статистика поретка, а да то не важи и за парно n.

Још један пример коришћења статистика поретка је екстремно одступање од узорачке средине $X_{(n)} - \bar{X}$. Узорачка средина се дефинише на следећи начин $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Све ове статистике имају важне примене. Екстреми се често јављају код статистичког проучавања поплава и суша, у теорији аукције, итд. Распон је познат по томе да обезбеђује добру оцену за стандардно одступање σ и има широку примену у области контроле квалитета. Екстремно одступање је основни алат за откривање аутлајера. Медијана је позната као робусна оцена локације.

1.2. Расподела статистике поретка

Нека обележје X има функцију расподелу F. Прост случајан узорак обима n из расподеле F је n-торка $(X_1, X_2, \dots, X_n)$, где су $X_1, X_2, \ . \ . \ . \ , X_n$ независне случајне величине са истом функцијом расподеле. Нека је $F_{(r)}(x)$, $r = 1, \dots, n$, функција расподеле r-те статистике поретка $X_{(r)}$. Функција расподеле максималне статистике поретка је

$$
\begin{aligned}
F_{(n)}(x) &= P\{X_{(n)} \leq x\} \\
&= P\{X_i \leq x\} = F^n(x).
\end{aligned}
$$

Слично, за минималну статистику поретка важи

$$
\begin{aligned}
F_{(1)}(x) &= P\{X_{(1)} \leq x\} = 1 - P\{X_{(1)} > x\} \\
&= 1 - P\{X_i > x\} = 1 - [1 - F(x)]^n.
\end{aligned}
$$

Одредимо функцију расподеле r -те статистике поретка $X_{(r)}$. Нека је $A_x = \{i : X_i \leq x\}$ и $|A_x|$ број елемената скупа A_x. Тада је

$$
\begin{aligned}
F_{(r)}(x) &= P\{X_{(r)} \leq x\} = P\{|A_x| \geq r\} \\
&= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i}. \\
&= \int_0^{F(x)} \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r} dt.
\end{aligned}
$$

Функција расподеле статистике поретка $X_{(r)}$ може се записати и у негативно биномној форми (Pinsker, Kipnis, Grechanovsky, 1986) на следећи начин

$$
F_{(r)}(x) = F^r \sum_{i=0}^{n-r} \binom{r+i-1}{r-1} [1 - F(x)]^i \qquad -\infty < x < +\infty
$$

Дати изрази важе за произвољну популацију, била она непрекидна или дискретна.

Одредимо сада густину расподеле статистике поретка $X_{(r)}$ за дискретан и непрекидан случај.

Како је густина $f(x) = F'(x)$ диференцирањем и елементарним трансформацијама добићемо густине расподела.

За дискретну популацију, густина расподеле статистике поретка $X_{(r)}$ добијена диференцирањем и елементарним трансформацијама функције расподеле је следећа

$$f_{(r)}(x) = P\{X_{(r)} = x\} = F_{(r)}(x) - F_{(r)}(x-1)$$

$$= \int_0^{F_{(r)}(x)} \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r}dt \; - \int_0^{F_{(r)}(x-1)} \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r}dt$$

$$= \int_{F_{(r)}(x-1)}^{F_{(r)}(x)} \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r}dt\,.$$

Специјално за $r = 1$ и $r = n$ добијамо

$$f_{(1)}(x) = [1 - F(x-1)]^n - [1 - F(x)]^n \;\text{ и }\; f_{(n)}(x) = [F(x)]^n - [F(x-1)]^n.$$

Ако је популација апсолутно непрекидна, онда је густина расподеле статистике поретка $X_{(r)}$ такође добијена диференцирањем функције расподеле следећа

$$f_{(r)}(x) = \frac{n!}{(r-1)!\,(n-r)!}\frac{d}{dx}\int_0^{F_{(r)}(x)} t^{r-1}(1-t)^{n-r}dt$$

$$= \frac{n!}{(r-1)!\,(n-r)!}[F(x)]^{r-1}[1-F(x)]^{n-r}f(x).$$

Специјално за $r = 1$ и $r = n$ добијамо

$$f_{(1)}(x) = n[1-F(x)]^{n-1}f(x), \quad f_{(n)}(x) = n[F(x)]^{n-1}f(x).$$

1.3. Заједничка расподела статистика поретка

Заједничка расподела статистика поретка се изводи на сличан начин. Природно, биће копликованија. Заједничка расподела статистика поретка $X_{(r)}$ и $X_{(s)}$ $(1 \le r < s \le n)$ је

$$F_{(r)(s)}(x,y) = F_{(s)}(y) \text{ за } x \ge y\,,$$

$$F_{(r)(s)}(x,y) = \sum_{j=s}^{n}\sum_{i=r}^{j} \frac{n!}{i!\,(j-i)!\,(n-j)!} F^i(x)[F(y)-F(x)]^{j-i}[1-F(y)]^{n-j} \text{ за } x \le y\,.$$

Дати изрази важе за сваку популацију била она непрекидна или дискретна.

Одредићемо заједничку густину расподеле за две и више статистика поретка за дискретан и непрекидан случај. Полазимо од дискретне расподеле.

Заједничка густина расподеле статистика поретка $X_{(r)}$ и $X_{(s)}$ $(1 \leq r < s \leq n)$ за дискретну популацију добија се диференцирањем њихове заједничке функције расподеле,

$$f_{(r)(s)}(x,y) = P(X_{(r)} = x, X_{(s)} = y)$$

$$= F_{(r)(s)}(x,y) - F_{(r)(s)}(x-1,y) - F_{(r)(s)}(x,y-1) + F_{(r)(s)}(x-1,y-1) \qquad x \leq y.$$

Овај израз се чини рачунски погоднији, док је интегрална форма (Khatri 1962) кориснија у теорији. Интегрални облик густине расподеле добићемо помоћу статистика поретка са униформном расподелом.

Нека је X случајна величина са расподелом F и нека је $F^{-1}(u) = \inf\{x : F(x) \geq u\}$ за $0 < u < 1$. Из непрекидности функције F следи да је $F\left(F^{-1}(u)\right) \geq u$ и $F^{-1}\left(F(x)\right) \leq x.$ Онда је $u \leq F(x)$ ако и само ако је $F^{-1}(u) \leq x$, (Serfling, 1980, p.3). Дакле, за $0 \leq F(x) \leq 1$,

$$P\{X \leq x\} = F(x) = P\{U \leq F(x)\} = P\{F^{-1}(U) \leq x\},$$

Што указује на то да X има исту расподелу као $F^{-1}(U)$, где U означава случајну величину са униформном расподелом. Одавде се одмах види да важи, (Scheffe, Tukey, 1945),

$$(X_{(1)}, X_{(2)}, \ldots, X_{(n)}) \triangleq (F^{-1}(U_{(1)}), F^{-1}(U_{(2)}), \ldots, F^{-1}(U_{(n)})),$$

где $\triangleq$ означава једнакост расподеле. Одавде добијамо уопштење заједничке густине расподеле у дискретном случају.

$$f_{(r)(s)}(x,y) = P\{F^{-1}\left(U_{(r)}\right) = x, F^{-1}\left(U_{(s)}\right) = y\}$$

$$= P\{F(x-1) < U_{(r)} \leq F(x), F(y-1) < U_{(s)} \leq F(y)\}$$

$$= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \iint v^{r-1}(w-v)^{s-r-1}(1-w)^{n-s}\, dvdw \quad x \leq y.$$

Област по којој интегралимо је $\{(u,v): v \leq w, F(x-1) \leq v \leq F(x), F(y-1) \leq w \leq F(y)\}$. Овај начин представљања заједничке густине расподеле две статистике поретка се примењује код других статистика поретка. Посебну улогу има у утврђивању одсуства Марковљевог својства код дискретних статистика поретка, што је приказано у следећем одељку.

У дискретном случају заједничка густина расподеле за $X_{(n_1)}, X_{(n_2)}, \ldots, X_{(n_k)}(1 \leq n_1 \ldots \leq n_k < n\,; 1 \leq k \leq n)$ је

$$f_{(n_1)\dots(n_k)}(x_1,\dots,x_k) = \frac{n!}{(n_1-1)!\,(n_2-n_1-1)!\cdots(n-n_k)!}\int\limits_B u_{n_1}^{n_1-1}$$

$$(u_{n_2}-u_{n_1})^{n_2-n_1-1}\cdots\left(u_{n_k}-u_{n_{k-1}}\right)^{n_k-n_{k-1}-1}\left(1-u_{n_k}\right)^{n-n_k}du_{n_k}\cdots du_{n_k}$$

где је B k-димензиони простор дат са $B=\left\{\left(u_{n_1},u_{n_2},\dots,u_{n_k}\right): u_{n_1}\le u_{n_2}\le\cdots\le u_{n_k}, F(x_r-1)\le u_r\le F(x_r) \text{ за } r=n_1,n_2,\dots,n_k\right\}$, (Balakrishnan, Nagaraja, 1992).

Сада прелазимо на непрекидан случај, где је заједничка густина расподеле статистика поретка $X_{(r)}$ и $X_{(s)}$ $(1\le r<s\le n)$ такође добијена диференцирањем заједничке функције расподеле, следећа

$$f_{(r)(s)}(x,y) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}F^{r-1}(x)f(x)[F(y)-F(x)]^{s-r-1}$$

$$f(y)[1-F(y)]^{n-s}.$$

Сада можемо добити заједничку расподелу за $X_{(n_1)}, X_{(n_2)},\dots,X_{(n_k)}$ $(1\le n_1\dots\le n_k< n\,;1\le k\le n)$

$$f_{(n_1)\dots(n_k)}(x_1,\dots,x_k) = \frac{n!}{(n_1-1)!\,(n_2-n_1-1)!\cdots(n-n_k)!}$$

$$F^{n_1-1}(x_1)f(x_1)[F(x_2)-F(x_1)]^{n_2-n_1-1}$$

$$f(x_2)\cdots[1-F(x_k)]^{n-n_k}f(x_k) \quad \text{за } x_1\le\ldots\le x_k.$$

Ако дефинишемо $x_1=-\infty, x_{\kappa+1}=+\infty, n_0=0, n_{k+1}=n+1$ претходни израз можемо да запишемо на следећи начин

$$n!\left[\prod_{j=1}^{k} f(x_j)\right]\prod_{j=0}^{k}\left\{\frac{\left[F\left(x_{j+1}\right)-F(x_j)\right]^{n_{j+1}-n_j-1}}{\left(n_{j+1}-n_j-1\right)!}\right\}.$$

Специјално, заједничка густина расподеле за свих n статистика поретка је

$$f_{(1)\dots(n)}(x_1,\dots,x_n) = n!\,f(x_1)\cdots f(x_n) \quad \text{за } x_1\le\ldots\le x_k\,.$$

1.4. Расподела распона

Помоћу заједничке функције расподеле k статистика поретка елементарним трансформацијама могу се добити функције расподеле за разне функције у којима фигуришу статистике поретка. На пример, за добијање функције расподеле распона статистика поретка $W_{rs} = X_{(s)} - X_{(r)}$ уводимо смену $w_{rs} = y - x$. Из заједничке густине расподеле две статистике поретка интеграљењем по x следи

$$f_{W_{rs}}(w_{rs}) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \int_{-\infty}^{+\infty} F^{r-1}(x) f(x) [F(x+w_{rs}) - F(x)]^{s-r-1}$$

$$\cdot f(x+w_{rs})[1 - F(x+w_{rs})]^{n-s} dx.$$

Специјално, за $r = 1,\ s = n$ и $W = X_{(n)} - X_{(1)}$ важи

$$f_W(w) = n(n-1) \int_{-\infty}^{+\infty} f(x)[F(x+w) - F(x)]^{n-2} f(x+w) dx\,.$$

Функција расподеле распона W је нешто једноставнија.

$$F_W(w) = n \int_{-\infty}^{+\infty} f(x) \int_{0}^{w} (n-1) f(x+w')[F(x+w') - F(x)]^{n-2} dw' dx$$

$$= n \int_{-\infty}^{+\infty} f(x)[F(x+w') - F(x)]^{n-1} \Big|_{w'=0}^{w'=w} dx$$

$$= n \int_{-\infty}^{+\infty} f(x)[F(x+w) - F(x)]^{n-1} dx\,.$$

Добијени резултат се може записати и на следећи начин

$$nf(x)dx[F(x+w) - F(x)]^{n-1}$$

и представља вроватноћу да вредност за неко X_i упада у интервал $(x, x+dx)$, а преосталих $n-1$ вредности упада у интервал $(x, x+w)$.

<u>Пример 1.1.</u> Одредимо расподелу статистика поретка као и статистике W_{rs} чија је популација униформна на интервалу $[0,1]$.

Због униформне расподеле статистике поретка ћемо означити са $U_{(1)}\,,\ U_{(2)}\,,\ \ldots\ ,\ U_{(n)}$.

Из густине расподеле статистике поретка $X_{(r)}$ добијамо

$$f_{(r)}(u) = \frac{n}{(r-1)!\,(n-r)!} u^{r-1}(1-u)^{n-r}, \text{за } 0 \le u \le 1$$

$$f_{(r)}(u) = 0\,, \text{иначе.}$$

Из заједничке густине расподеле статистика поретка $X_{(r)}$ и $X_{(s)}$ добијамо

$$f_{(r)(s)}(u,v) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s}, 0 \le u \le v \le 1$$

$$f_{(r)(s)}(u,v) = 0 \text{, иначе.}$$

Како је $f(u + w_{rs}) = 0$ за $u > 1 - w_{rs}$ из функције густине распона W_{rs} добијамо

$$f_{W_{rs}}(w_{rs}) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \int_0^{1-w_{rs}} u^{r-1} {w_{rs}}^{s-r-1}(1-u-w_{rs})^{n-s} du \,.$$

Увођењем смене $u = v(1 - w_{rs})$ добијамо

$$f_{W_{rs}}(w_{rs}) = \frac{(n+1)!}{(s-r)!\,(n-s+r+1)!} {w_{rs}}^{s-r-1}(1-w_{rs})^{n-s+r} \text{ за } 0 \le w_{rs} \le 1 \,.$$

Видимо да статистика W_{rs} има бета расподелу која зависи само од $s - r$, а не од s и r појединачно.

1.5. Условна расподела, статистике поретка као ланац Маркова

Из заједничке расподеле за n статистика поретка следи да је условна густина расподеле за $X_{(r+1)}, X_{(r+2)}, \dots, X_{(s-1)}$ где је $X_{(i)} = x_i$ за $i \le r$ и $i \ge s$ где је $1 \le r < s \le n$, следећа

$$f_{X_{(r+1)},X_{(r+2)},\dots,X_{(s-1)}|X_{(i)}=x_i\,,\ i\le r, i\ge s}(x_{r+1},\dots,x_{s-1})$$

$$= (s-r-1)! \prod_{j=r+1}^{s-1} \frac{f(x_j)}{F(x_s) - F(x_r)} \qquad x_1 < \dots < x_n \,.$$

Теорема 1.1. За случајан узорак обима n из непрекидне расподеле, условна расподела за $X_{(r+1)}, X_{(r+2)}, \dots, X_{(s-1)}$, где је $X_{(i)} = x_i$ за $i \le r$ и $i \ge s$, $r < s$, је само расподела свих статистика поретка из узорка обима $s - r - 1$ добијена из $f(x_j)/[F(x_s) - F(x_r)]$, $(x_1 < x < x_s)$.

Како је условна расподела дата без x_i за $i < r$ и $i > s$, онда $X_{(r+1)}, X_{(r+2)}, \dots, X_{(s-1)}$ не зависе од $X_{(1)}, X_{(2)}, \dots, X_{(r-1)}$ и $X_{(s+1)}$, $X_{(s+2)}, \dots, X_{(n)}$ када су $X_{(r)}$ и $X_{(s)}$ дате. Након одређивања условне густине расподеле за статистике поретка нижег ранга закључујемо да важи следеће

$$f_{X_{(r+1)},X_{(r+2)},\dots,X_{(n)}|X_{(i)}=x_i\,,\dots,X_{(r)}=x_r}(x_{r+1},\dots,x_n)$$

$$= f_{X_{(r+1)},X_{(r+2)},\dots,X_{(n)}|\,X_{(r)}=x_r}(x_{r+1},\dots,x_n),$$

на основу чега утврђујемо да статистике поретка узорка из непрекидне расподеле формирају ланац Маркова. Следећи израз даје густину прелаза.

$$f_{X_{(r+1)}|X_{(r)}=x_r}(y) = (n-r)\left\{\frac{1-F(y)}{1-F(x)}\right\}^{n-r-1} \frac{f(y)}{1-F(x)} \quad y > x\,,$$

Нека $z_{(1)} \leq \cdots \leq z_{(n)}$ означавају статистике поретка из узорка обима n из експоненцијалне расподеле

$$f(z) = e^{-x} \quad 0 \leq z < \infty,$$

онда је заједничка густина расподеле за $Z_{(i)}$ једнака

$$n!\exp\left(\sum_{i=1}^{n} z_{(i)}\right), \qquad 0 \leq z_{(1)} \leq \cdots \leq z_{(n)} < \infty,$$

што се може записати и на следећи начин (Sukhatme, 1937),

$$n!\exp\left[-\sum_{i=1}^{n}(n-i-1)\big(z_{(i)} - z_{(i-1)}\big)\right],$$

где је $z_{(0)} = 0$.

Уводимо смену $y_i = (n-i+1)\big(z_{(i)} - z_{(i-1)}\big)\; i = 1, \dots, n, y_i \epsilon (0,\infty)$. Y_i су независне величине са експоненцијалном расподелом. Сада $Z_{(r)}$ можемо да изразимо на следећи начин

$$Z_{(r)} = \sum_{i=1}^{r}(Z_{(i)} - Z_{(i-1)}) = \sum_{i=1}^{r}\frac{Y_i}{n-i+1}.$$

Ово је линеарна функција независних експоненцијалних промењљивих. Из чега следи да $Z_{(1)}, \dots, Z_{(n)}$ формирају адитивни ланац Маркова (Renyi, 1953).

Сменом $u = F(x)$ кад је у питању апсолутно непекидна расподела, преводимо $X_{(i)}$ у $U_{(i)}$ $(i = 1, \dots, n)$, статистику поретка са униформном расподелом $U(0,1)$. Како је $z = -\log u$ монотоно опадајућа функција по u, а $-\log U$ има експоненцијалну расподелу, следи је

$$Z_{(r)} = -\log U_{(n-r+1)} \qquad r = 1, \dots, n\,.$$

Из предходног израза и трансформације $y_i = (n-i+1)\big(z_{(i)} - z_{(i-1)}\big)$ следи да су količници

$$\frac{U_{(r)}}{U_{(r+1)}} = \exp\left(\frac{-Y_{n-r+1}}{r}\right)$$

међусобно независни ($r = 1, \dots, n$; $U_{(n+1)} = 1$). Одатле следи и да су

$$\left(\frac{U_{(r)}}{U_{(r+1)}}\right)^r = \exp(-Y_{n-r+1})$$

међусобно независне величине са униформном расподелом $U(0,1)$, (Malamqist, 1950).За Паретову расподелу на интервалу $(0,1)$, $X_{(r)}/X_{(r+1)}$ је независно, а на интервалу $(1,\infty)$, $X_{(r+1)}/X_{(r)}$ је независно.

Марковљево својство не важи у случају дискретне расподеле ако нису испуњени одређени услови. Заправо показаћемо да важи следеће (Nagaraja, 1982)

$$f_{X_{(i+1)}|X_{(i-1)}=x,\ X_{(i)}=y}(z) < f_{X_{(i+1)}|\ X_{(i)}=y}(z) \quad \text{за } x < y < z\,.$$

Због једноставности, претпоставићемо да је $f(x)$ дискретна над целим бројевима.

$$f_{X_{(i+1)}|X_{(i-1)}=x,\ X_{(i)}=y}(z) = \frac{f_{(i-1)(i)(i+1)}(x,y,z)}{f_{(i-1)(i)}(x,y)} =$$

$$= (n-i)[F(y)-F(y-1)]\frac{\int_{F(x-1)}^{F(x)}(1-w)^{n-i-1}dw}{\int_{F(y-1)}^{F(y)}(1-v)^{n-i}dv}$$

$$= (n-i)\frac{\int_{F(y-1)}^{F(y)} v^{i-1}dv \int_{F(x-1)}^{F(x)}(1-w)^{n-i-1}dw}{\int_{F(y-1)}^{F(y)} v^{i-1}\,(1-v)^{n-i}dv}$$

$$\frac{1}{b-a}\int_a^b v^{i-1}\,(1-v)^{n-i}dv < \left[\frac{1}{b-a}\int_a^b v^{i-1}\,dv\right]\left[\frac{1}{b-a}\int_a^b (1-v)^{n-i}dv\right].$$

Где је $0 \le a = F(y-1) < F(y) = b \le 1$. Ова неједнакост следи из чињенице да када је случајна величина Y униформна на интервалу $[a,b]$, Y^{i-1} и $(1-Y)^{n-i}$ су негативно корелисане. Ова неједнакост такође важи и кад је $x = y = z$, а обрнута кад је $x = y < z$ или $x < y = z$.

1.6. Сродне статистике

Нека је $X_1, X_2, \dots$ бесконачан низ независних, равномерно распоређених променљивих са расподелом $F(x)$ и густином $f(x)$. За $i \ge 2$, X_i зовемо горња рекордна вредност низа ако је $X_i = \max\{X_1, X_2, \dots, X_i\}$. По договору, X_1 је горња рекордна вредност. Индекси за које се јављају ове граничне вредности представљају низ случајних променљивих $T_1, T_2, \dots$ за које важи $T_1 = 1, T_n = \min\{i | X_i > X_{T_{n-1}}\}$. Горње рекордне вредности X_{T_i}

представљају различите елементе у низу узастопних максимума од X_i. Слично се дефинишу доње рекордне вредности.

Заједничка расподела за $T_1, T_2, \dots, T_n$ и $X_{T_1}, X_{T_2}, \dots, X_{T_n}$ је

$$\prod_{i=1}^{n} f(x_i) \prod_{j=1}^{n-1} \left[F(x_j)\right]^{t_{j+1}-t_j}$$

за $x_1 < \cdots < x_n$ и $1 = t_1 < \cdots < t_n$. А заједничка густина расподеле за $X_{T_1}, X_{T_2}, \dots, X_{T_n}$

$$f_{1\dots n}(x_1, \dots, x_n) = f(x_n) \prod_{j=1}^{n-1} r\big(x_j\big), \quad x_1 < \cdots < x_n,$$

где је $r(x) = f(x)/[1 - F(x)]$ стопа ризика или стопа неуспеха функције (Arnold, Nagaraja, Balakrishnan, 1998).

Stigler (1977) је увео фракционе статистике поретка. За $\lambda > 0, k \geq 1,$ и $0 = t_0 < t_1 < \cdots < t_k < 1$ где случајне величине $U(t_1), \dots, U(t_k)$ имају Дирихлеову заједничку расподелу са густином

$$f(u_1, \dots, u_k) = \frac{\Gamma(\lambda)}{\Gamma(\lambda t_1)\Gamma(\lambda(t_2 - t_1)) \dots \Gamma(\lambda(1 - t_k))}$$

$$\cdot \prod_{i=1}^{k} (u_i - u_{i-1})^{\lambda(t_i - t_{i-1}) - 1} (1 - u_k)^{\lambda(1-t_k)-1}$$

$$0 = u_0 < u_1 < \cdots < u_k < 1.$$

За $\lambda = n + 1$, $U(t)$ $(0 < t < 1)$ је $\beta((n+1)t, (n+1)(1-t))$ променљива, а за $t = r/(n+1)$, понаша се као $U_{(r)}$. Тада, $U(t) = U_{(\overline{n+1}\, t)}$, можемо посматрати као униформну фракциону статистику поретка, а за узорак из произвољне расподеле, $X_{(\overline{n+1}\, t)}$ можемо представити као $F^{-1}\big(U_{(\overline{n+1}\, t)}\big)$. Фракциона статистика поретка је случајна мешавина обичних статистика поретка које задовољавају услов, (Jones, 2002),

$$U_{(\overline{n+1}\, t)} \triangleq V \cdot U_{(r+1)} + (1 - V) \cdot U_{(r)},$$

Где је $r = [(n+1)t]$, $c = (n+1)t - r$, а V је независна $\beta(c, 1-c)$ променљива. Ако је $c = 0$ онда је $V \equiv 0$.

Дефинишимо уопшту униформну статистку поретка коју је увео Kamps (1995). Нека важи $X \triangleq F^{-1}(U)$, што је већ дефинисано у претходном одељку. За $n \geq 2$, нека је $\widetilde{m} = (m_1, \dots, m_{n-1})$ где су $m_1, \dots, m_{n-1}$ реалне вредности, параметри такви да је $\gamma_i = k + n - i + \sum_{j=i}^{n-1} m_j \geq 1$ за $1 \leq i \leq (n-1)$ и $k \geq 1$. Уопштена униформна статистика поретка $U(i, n, \widetilde{m}, k)$, $i = 1, \dots, n$, има заједничку густину расподле

$$f(u_1, \dots, u_n) = k\left(\prod_{j=1}^{n-1} \gamma_j\right)\left(\prod_{i=1}^{n-1} (1 - u_i)^{m_i}\right)(1 - u_n)^{k-1},$$

где је $0 < u_1 < \cdots < u_n < 1$. Ове случајне промењиве показују Марковљево својство са вероатноћама прелаза

$$P\{U(i, n, \widetilde{m}, k) > v | U(i - 1, n, \widetilde{m}, k) = u\} = \left\{\frac{1 - v}{1 - u}\right\}^{\gamma_i},$$

где је $0 \leq u < v \leq 1$.

1.7. Непараметарски интервали поверења за квантиле

Квантил реда p је вредност која је већа или једнака од p% елемената из узорка. Специјално, 25%-тни квартил се назива први квантил и означава се са $\xi_{1/4}$, 50%-тни $\xi_{1/2}$ је медијана узорка, а 75%-тни $\xi_{3/4}$ назива се трећи квартил. Разлика првог и трећег квартила се назива квартилни распон.

Нека је X случајна величина са непрекдном расподелом $F(x)$. Тада једначина

$$F(x) = p,\ 0 < p < 1$$

има јединствено решење $x = \xi_p$, што називамо квантил реда p. Ако $F(x)$ није строго растућа онда једначина $F(x) = p$ има решење у неком интервалу, и било која тачка из тог интервала може да буде квантил реда p.

Ако X има дискретну расподелу онда важи

$$P\{X < \xi_p\} \leq p \leq P\{X \leq \xi_p\}.$$

Уопштено, имамо да је $\xi_p = F^{-1}(p)$, где је $F^{-1}(p) = \inf\{x: F(x) \geq p\}$.

Вероватноћа да ξ_p припада интервалу $\left(X_{(r)}, X_{(s)}\right)$ зависи од r, s, n и p, а не и од $F(x)$. Вроватноћу догађаја $X_{(r)} \leq \xi_p$, где је X има непрекидну или дискретну расподелу, можемо изразити на следећи начин

$$P\{X_{(r)} \leq \xi_p\} = P\{X_{(r)} \leq \xi_p \leq X_{(s)}\} + P\{X_{(s)} < \xi_p\}$$

$$P\{X_{(r)} \leq \xi_p \leq X_{(s)}\} = P\{X_{(r)} \leq \xi_p\} - P\{X_{(s)} < \xi_p\}.$$

За непрекидно X вероватноћа догађаја да ξ_p припада интервалу $\left(X_{(r)}, X_{(s)}\right)$ коју ћемо означизи са $\pi(r, s, n, p)$ је

$$\pi(r,s,n,p) = \int\limits_0^p \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r}dt - \int\limits_0^p \frac{n!}{(s-1)!\,(n-s)!} t^{s-1}(1-t)^{n-s}dt$$

$$= \sum_{i=1}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}.$$

Добијене резултате дао је Thompson (1936).

А у дискретном случају из вероватноћа $P\{X < \xi_p\} \leq p$ и $P\{X \leq \xi_p\} \geq p$ следи

$$P\{X_{(r)} \leq \xi_p\} \geq \int\limits_0^p \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r}dt$$

$$P\{X_{(s)} < \xi_p\} \leq \int\limits_0^p \frac{n!}{(s-1)!\,(n-s)!} t^{s-1}(1-t)^{n-s}dt,$$

Како је вероватноћа догађаја $X_{(r)} \leq \xi_p \leq X_{(s)}$ једнака разлици вероватноћа догађаја $X_{(r)} \leq \xi_p$ и $X_{(s)} \leq \xi_p$ важи да је

$$P\{X_{(r)} \leq \xi_p \leq X_{(s)}\} \geq \pi(r,s,n,p).$$

Слично, можемо добити и

$$P\{X_{(r)} < \xi_p < X_{(s)}\} \leq \pi(r,s,n,p).$$

Лева страна ових неједнакости не зависи од $F(x)$, њихова горња и доња ограничења не зависе од расподеле (Scheffe, Tukey, 1945).

Ако је интервал поверења са коефицијентом који је већи или једнак од $1-\alpha$ за дато n и p и произвољне r и s онда је и $\pi \geq 1-\alpha$. Одабир r и s је донекле прозвољан, најбоље је узети да $s-r$ буде што мање. Ако је $p = 1/2$ узећемо да је $s = n-r+1$, у том случају ће π бити

$$\pi\left(r, n-r+1, n, \frac{1}{2}\right) = 2\int\limits_0^{1/2} \frac{n!}{(r-1)!\,(n-r)!} t^{r-1}(1-t)^{n-r}dt - 1 = 2^{-n}\sum_{i=r}^{n-r}\binom{n}{i}.$$

Интервали поверења за разлике квантила $\xi_q - \xi_p$ $q > p$ могу се одредити помоћу разлика статистика поретка $X_{(s)} - X_{(r)}$. Показаћемо да важи, (Chu, 1957),

$$P\{X_{(s)} - X_{(r)} \geq \xi_q - \xi_p\} \geq I_p(r, n-r+1) - I_q(s, n-s+1) = L$$

$$P\{X_{(v)} - X_{(u)} \leq \xi_q - \xi_p\} \geq I_q(v, n-v+1) - I_p(u, n-u+1) = L'.$$

Где је $I_p(r, n-r+1) = \int_0^p \frac{n!}{(r-1)!(n-r)!} t^{r-1}(1-t)^{n-r} dt$.

Доказ. $P\{X_{(s)} - X_{(r)} \geq \xi_q - \xi_p\} \geq P\{X_{(s)} \geq \xi_q,\ X_{(r)} \leq\ \xi_p\}$

$$\geq P\{X_{(s)} \geq \xi_q\} + P\{\ X_{(r)} \leq\ \xi_p\} - 1$$

$$= P\{\ X_{(r)} \leq\ \xi_p\} - P\{X_{(s)} \geq \xi_q\}$$

$$\geq I_p(r, n-r+1) - I_q(s, n-s+1)$$

Слично се добија и друга неједнакост. □

За било које $\alpha \in (0,1)$ и довољно велико n постоји бар један скуп вредности $\{r, s, u, v\}$ за које важи да је $L \geq 1-\alpha$ и $L' \geq 1-\alpha$. $X_{(s)} - X_{(r)}$ и $X_{(v)} - X_{(u)}$ су горња и доња граница поверења за $\xi_q - \xi_p$ са коефицијентом поверења $\geq 1-\alpha$. У случају $q = 1-p$, где је $s = n-r+1$ и $v =\ n-u+1$ добијамо

$$I_p(r, n-r+1) \geq 1 - \frac{1}{2}\alpha, \qquad I_p(u, n-u+1) \leq\ \frac{1}{2}\alpha.$$

У претходном доказу неједнакости $\{X_{(s)} \geq \xi_q,\ X_{(r)} \leq\ \xi_p\}$ се могу записати и као $\{\ X_{(r)} \leq\ \xi_p < \xi_q \leq X_{(s)}\}$ и одатле важи

$$P\{\ X_{(r)} \leq\ \xi_p < \xi_q \leq X_{(s)}\} \geq L,$$

$$P\{\xi_p \leq X_{(u)} \leq X_{(v)} \leq \xi_q\} \geq L'.$$

Дакле, интервале $[\ X_{(r)}, X_{(s)}]$ и $[X_{(u)}, X_{(v)}]$ можемо назвати горњи и доњи интервал поверења за интервал квантила $[\xi_p, \xi_q]$, (Wilks, 1962).

2. Очекиване вредности и моменти статистика поретка

2.1. Основне формуле

У даљем раду позабавићемо се моментима статистика поретка, поготово очекивањем, варијансом и коваријансом. Линеарне функције статистика поретка, су корисне у оцењивању параметара. Познавање очекивања, варијансе и коваријансе омогућава нам да пронађемо очекиване вредности и варијансу линеарне функције, и одатле нам помажу да нађемо оцене и њихову ефикасност.

Понекад ће бити згодно да истакнемо обим узорка, па ћемо за $X_{(r)}$ користити ознаку $X_{r:n}$, $r = 1, \dots, n$. Математичко очекивање $E(X_{r:n})$ ћемо означиити са $\mu_{r:n}$. За непрекидну расподелу важи

$$\mu_{r:n} = \int_{-\infty}^{\infty} x f_{r:n}(x) dx$$

или

$$\mu_{r:n} = \frac{n!}{(r-1)!\,(n-r)!} \int_{-\infty}^{\infty} x F^{r-1}(x)[1 - F(x)]^{n-r} f(x) dx.$$

Увођењем смене $u = F(x)$, математичко очекивање можемо да запишемо и на следећи начин

$$\mu_{r:n} = \frac{n!}{(r-1)!\,(n-r)!} \int_{0}^{1} F^{-1}(u) u^{r-1} (1-u)^{n-r} du,$$

где је $F^{-1}(u)$ инверз функције $F(x)$,

$$E(X_{r:n}) = E[F^{-1}(U_{r:n})].$$

Како је $0 \leq F(x) \leq 1$, из неједнакости

$$|\,\mu_{r:n}| \leq \frac{n!}{(r-1)!\,(n-r)!} \int_{-\infty}^{\infty} |x| f(x) dx$$

$$= \frac{n!}{(r-1)!\,(n-r)!} E|X|$$

видимо да $\mu_{r:n}$ постоји ако постоји $E(X)$, пошто на основу особина математичког очекивања знамо да $E(X)$ и $E|X|$ истовремено постоје или не постоје. Обрнуто не мора да важи. У том смислу можемо да кажемо да ако постоји $E[g(x)]$, где је $g(x)$ нека функција од x, онда постоји и $E[g(X_{r:n})]$. Специјални случајеви су $g(x) = x^k, (x - \mu_{r:n})^k$, и e^{tx}, дају редне моменте, централне моменте и моменте генераторних функција од $X_{r:n}$. Моменат k-тог реда је

$$\mu_{r:n}^{(k)} = \int_{-\infty}^{\infty} x^k f_{r:n}(x)dx$$

$$= \frac{n!}{(r-1)!\,(n-r)!} \int_{-\infty}^{\infty} x^k F^{r-1}(x)[1-F(x)]^{n-r} f(x)dx$$

$$= E(X_{r:n}^k).$$

Sen (1959) је показао да, ако постоји $E|X|^\delta$ за неко $\delta > 0$, онда $\mu_{r:n}^{(k)}$ постоји за свако r за које је $r_0 < r < n - r_0 - 1$, где је $r_0\delta = k$.

Слично можемо да дефинишемо моменте производа статистика поретка

$$\mu_{r,s:n} = \iint_{x<y} xy f_{r,s}(x,y)dxdy$$

$$= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \iint_{x<y} xyF^{r-1}(x)[F(y)-F(x)]^{s-r-1}$$
$$\times [1-F(x)]^{n-s} f(x)f(y)dxdy$$

$$= E(X_{r:n}X_{s:n}) \quad r,s = 1, \dots, n.$$

Коваријанса статистика поретка $X_{r:n}$и $X_{s:n}$, која се означава $\mathrm{Cov}(X_{r:n}X_{s:n}) = \sigma_{r,s:n}$ је

$$\sigma_{r,s:n} = E[(X_{r:n} - \mu_{r:n})(X_{s:n} - \mu_{s:n})],$$

или

$$\sigma_{r,s:n} = \mu_{r,s:n} - \mu_{r:n}\,\mu_{s:n}$$
$$= E(X_{r:n}X_{s:n}) - E(X_{r:n})E(X_{s:n}).$$

Важи као и обично да је $\sigma_{r,s:n} = \sigma_{s,r:n}$, и да је $\sigma_{r,r:n}$ односно $\sigma_{r:n}^2$ варијанса (дисперзија) случајне величине $X_{r:n}$, која се може означити и као $D(X_{r:n})$, и једнака је

$$\sigma_{r:n}^2 = \int_{-\infty}^{+\infty} (x - \mu_{r:n})^2 f_{r:n}(x)dx\,,$$

Док је за $r < s$ коваријанса једнака

$$\sigma_{r,s:n} = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} (x - \mu_{r:n})(y - \mu_{s:n}) f_{r,s:n}(x,y)dxdy,$$

где је $f_{r,s:n}(x,y)$ заједничка густина расподеле за $X_{r:n}$ и $X_{s:n}$.

Опет, увођењем смене, за било коју расподелу $F(x,y)$ имамо да је

$$\sigma_{r,s:n} = E\{[F^{-1}(U_{r:n}) - \mu_{r:n}][F^{-1}(U_{s:n}) - \mu_{s:n}]\}$$

$$= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \int_0^1 \int_0^v [F^{-1}(u) - \mu_{r:n}][F^{-1}(v) - \mu_{s:n}]\, u^{r-1}(v - u)^{s-r-1} \cdot (1-v)^{n-s} dudv,$$

<u>Пример 2.1.</u> За $f(x)$ униформно на $[0,1]$ добијамо

$$\mu_{r:n} = \frac{n!}{(r-1)!\,(n-r)!} \int_{-\infty}^{\infty} x \cdot x^{r-1}(1-x)^{n-r} dx$$

$$= \frac{n!}{(r-1)!\,(n-r)!} \Big/ \frac{(n+1)!}{r!\,(n-r)!}$$

$$= \frac{r}{n+1}.$$

Овај резултат указује на на то да статистике поретка деле област ипод криве $y = f(x)$ на $n+1$ делова, сваки са очекиваном вредношћу $1/(n+1)$.

Општи приступ за процену производа момената приказали су David и Johnson (1954) помоћу четири промнљиве

$$f_{r,s,t,u:n}(x_1,x_2,x_3,x_4) = \frac{n!}{(r-1)!\,(s-r-1)!\,(t-s-1)!\,(u-t-1)!\,(n-u)!} \cdot x_1^{r-1}(x_2-x_1)^{s-r-1}(x_3-x_2)^{t-s-1}(x_4-x_3)^{u-t-1}(1-x_4)^{n-u},$$

за $1 \le r < s < t < u \le n$ и $0 \le x_1 \le x_2 \le x_3 \le x_4 \le 1$, уводимо смене $x_4 = y_4$, $x_3 = y_3y_4$, $x_2 = y_2y_3y_4$, $x_1 = y_1y_2y_3y_4$.

Означићемо константу са C, и имајући у виду да је Јакобијан $y_2y_3^2y_4^3$ добијамо

$$f(y_1,y_2,y_3,y_4) = Cy_1^{r-1}(1-y_1)^{s-r-1}y_2^{s-1}(1-y_2)^{t-s-1}y_3^{t-1}(1-y_3)^{u-t-1}y_4^{u-1}(1-y_4)^{n-u}$$

$$0 \le y_i \le 1, i = 1, \dots, 4.$$

На осову овога видимо да су Y_i, а отуда и количници $X_{r:n}/X_{s:n}$, $X_{s:n}/X_{t:n}$, $X_{t:n}/X_{u:n}$ и $X_{u:n}$ независни. Дакле, имамо

$$E(X_{r:n}^a X_{s:n}^b X_{t:n}^c X_{u:n}^d) = \frac{1}{B(r,s-r)} \int_0^1 y_1^{r-1+a}(1-y_1)^{s-r-1} dy_1 \dots$$

$$\times \frac{1}{B(u,n-u+1)} \int_0^1 y_4^{u-1+a+b+c+d}(1-y_4)^{n-u} dy_4$$

$$= \frac{(r-1+a)!\,(s-1+a+b)!\,(t-1+a+b+c)!\,(u-1+a+b+c+d)!\,n!}{(r-1)!\,(s-1+a)!\,(t-1+a+b)!\,(u-1+a+b+c)!\,(n+a+b+c+d)!}.$$

Генерално, за статистике поретка $X_{r_i:n}(i=1,\dots,k)$ важи, (F. N. David, Johnson, 1954),

$$E\left(\prod_{i=1}^{k} X_{r_{i:n}}^{a_i}\right) = \frac{n!}{\left(n+\sum_{i=1}^{k} a_i\right)!} \prod_{i=1}^{k} \frac{\left(r_i - 1 + \sum_{j=1}^{i} a_j\right)!}{\left(r_i - 1 + \sum_{j=1}^{i-1} a_j\right)!}.$$

Дакле, ако је $p_r = r/(n+1)$, $q_r = 1 - p_r$, специјално за $r \le s \le t$ закључујемо да важи

$$\mu_{r:n} = p_r, \quad \sigma_{r,s:n} = \frac{p_r q_s}{n+2}$$

$$E[(X_{r:n} - \mu_{r:n})(X_{s:n} - \mu_{s:n})(X_{t:n} - \mu_{t:n})] = \frac{2p_r(q_s - p_s)q_t}{(n+2)(n+3)},$$

и

$$E(X_{r:n} - \mu_{r:n})^4 = \frac{3p_r^2 q_r^2}{(n+2)^2} + \frac{6p_r q_r}{(n+2)(n+3)(n+4)}\left[(q_r - p_r)^2 - \frac{n+3}{n+2} p_r q_r\right].$$

Поред момената линеарних функција статистика поретка, можемо да одредимо и моменте распона:

$$E(W) = \mu_{n:n} - \mu_{1:n},$$

$$D(W) = \sigma_{n:n}^2 - 2\sigma_{1,n:n} + \sigma_{1:n}^2,$$

што се у случају симетричности у односу на $x = 0$ своди на

$$E(W) = 2\mu_{n:n}$$

$$D(W) = 2(\sigma_{n:n}^2 - 2\sigma_{1,n:n})$$

Пример 2.2. За униформну расподелу на интервалу [0,1] имамо

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

Што може да се провери помоћу густине распона за $r = 1$ и $s = n$,

$$f_{W_n}(w) = \frac{1}{B(n-1,2)} w^{n-2}(1-w) \qquad 0 \le w < 1,$$

одакле добијамо

$$D(W_n) = \frac{(n-1)n}{(n+1)(n+2)} - \left(\frac{n-1}{n+1}\right)^2 = \frac{2(n-1)}{(n+2)(n+1)^2}.$$

Алтернативна формула за $\boldsymbol{\mu_{r:n}}$

Лема 2.1. Ако је X случајна величина са функцијом расподеле $F(x)$, онда за $E(|X|^\delta) < \infty$ и $\delta > 0$ следи,

$$\lim_{x\to\infty} x^\delta[1-F(x)] = 0 \ \text{ и } \lim_{x\to-\infty} x^\delta F(x) = 0.$$

Доказ. За $x \ge 0$,

$$x^\delta[1-F(x)] = x^\delta \int_x^\infty dF(t) \le \int_x^\infty t^\delta dF(t).$$

Али како је $E(|X|^\delta) < \infty$, десна страна једнакости тежи нули кад $x \to \infty$. Слично

$$\lim_{x\to-\infty} |x|^\delta F(x) = 0. \qquad \square$$

Ако је $E(X) < \infty$ можемо да пишемо

$$E(X) = \int_{-\infty}^{0} x dF(x) - \int_0^\infty x d[1-F(x)]$$

и из претходне леме важи

$$E(X) = \int_0^\infty [1-F(x)]dx - \int_{-\infty}^{0} F(x)dx$$

$$= \int_0^\infty [1-F(x)-F(-x)]dx.$$

Ови резултати нам дају алтернативну формулу за $\mu_{r:n}$ ако се $F(x)$ замени са $F_{r:n}(x)$.

$$\mu_{r:n} = \int_0^\infty [1-F_{r:n}(x)-F_{r:n}(-x)]dx.$$

Поставимо да је $r = n$ и $r = 1$, из претходног израза добијамо основну формулу, (Tippett, 1925; Cox, 1954),

$$E(W) = \int_{-\infty}^{\infty} \{1 - F^n(x) - [1 - F(x)]^n\} dx.$$

Такође кад је X симетрично у односу на 0 имамо

$$\mu_{r:n} = \int_{0}^{\infty} [F_{n-r+1:n}(x) - F_{r:n}(x)] dx.$$

Основне релације

Општи начини провере прорачуна могу се добити коришћењем релације

$$\left(\sum_{r=1}^{n} X_{r:n}^{k}\right)^{m} = \left(\sum_{r=1}^{n} X_{r}^{k}\right)^{m}.$$

Помоћу ове једнакости можемо добити неколико идентитета за појединачне моменте и моменте производа статистика поретка. На пример ако узмемо да су (k, m) редом $(1,1), (2,1)$ и применимо очекивање са обе стране добијамо следеће једнакости

$$\sum_{r=1}^{n} \mu_{r:n} = nE(X) = n\mu$$

$$\sum_{r=1}^{n} E(X_{r:n}^{2}) = nE(X^2).$$

На сличан начин, за $k = 1$ и $m = 2$ имамо

$$\sum_{r=1}^{n} X_{r:n}^{2} + 2\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} X_{r:n} X_{s:n} = \sum_{r=1}^{n} X_{r}^{2} + 2\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} X_{r} X_{s}.$$

Применом очекивања са обе стране једнакости, добијамо

$$\sum_{r=1}^{n} \mu_{r:n}^{2} + 2\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \mu_{r,s:n} = nE(X^2) + \frac{2n(n-1)}{2} [E(X)]^2.$$

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \mu_{r,s:n} = \frac{n(n-1)}{2} [E(X)]^2$$

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} E(X_{r:n} X_{s:n}) = \frac{n(n-1)}{2} \mu^2.$$

Квадрирањем једнакости $\sum (X_{r:n} - \mu_{r:n}) = \sum (X_r - \mu)$ имамо

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\sigma_{r,s:n} = n\sigma^2.$$

Добијени резултати су прменљиви и у дискретном и у непрекидном случају.

Тврђење 2.1. (Joshi, Balakrishnan, 1982) За $\mu = 0$, важи

$$\sum_{s=r+1}^{n}\mu_{r,s:n} + \sum_{i=1}^{r}\mu_{r+1,i:n} = 0, \quad r = 1, \dots, n-1,$$

Ако су $\mu_{r,s:n}$ уређени у виду матрице, $n-1$ сума момената производа који се налазе између узастопних момената $\mu_{r,r:n}$ су нуле.

Доказ. Нека је

$$J_1 = n\frac{n!}{(r-1)!\,(n-r)!}\int_0^1\int_0^v F^{-1}(u)F^{-1}(v)u^{r-1}(1-u)^{n-1-r}dudv.$$

Уместо $1-u$ писаћемо $(1-v)+(v-u)$

$$J_1 = n\frac{n!}{(r-1)!\,(n-r)!}\sum_{j=0}^{n-1-r}\binom{n-1-r}{j}\int_0^1\int_0^v F^{-1}(u)F^{-1}(v)u^{r-1}(v-u)^j$$

$$\cdot\,(1-v)^{n-1-r-j}dudv$$

$$= \frac{n!}{(r-1)!\,(n-r-1)!}\sum_{j=0}^{n-1-r}\binom{n-1-r}{j}\frac{(r-1)!\,j!\,(n-1-r-j)!}{n!}\mu_{r,r+j+1:n}$$

$$= \sum_{s=r+1}^{n}\mu_{r,s:n}, \qquad s = r+j+1.$$

Слично, након смене $u = v + (u-v)$ добијаћемо да је

$$J_2 = n\frac{(n-1)!}{(r-1)!\,(n-r-1)!}\int_0^1\int_v^1 F^{-1}(u)F^{-1}(v)u^{r-1}(1-u)^{n-1-r}dudv,$$

$$= \sum_{i=1}^{r}\mu_{i,r+1:n},$$

$$J_1 + J_2 = n\mu_{r:n-1}\mu = 0. \qquad \square$$

2.2. Моменти у случају непрекидне расподеле

Моменти статистика поретка нижег ранга у случајном узорку обима n могу се експлицитно одредити само за неке једноставније популације, као што су популације са униформном или експоненцијалном расподелом. Обично је потребна нумерчка интеграција, која је често захтевна, али сада су доступне таблице очекивања, варијанси и коваријанси за многе стандардне расподеле узорка.

За неке специјалне расподеле могуће је користити релације које смо приказали у претходном одељку. Поред тога што обезбеђују проверу прорачуна, могу и да доведу до неких поједностављења. У овом одељку ћемо детљније размотрити случај нормалне расподеле.

Густина и функција нормалне расподеле су дате

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty < x < \infty$$

$$F(x) = \Phi(x) = \int_{-\infty}^{x} \phi(t) dt.$$

Важи следеће

$$\sum_{s=1}^{n} \mu_{r,s:n} = 1\ , \quad \sum_{s=1}^{n} \sigma_{r,s:n} = 1, \quad r = 1, \dots, n.$$

Доказ. Из независности $X_{r:n} - \bar{X}$ и $\bar{X}$ следи да је

$$E[(X_{r:n} - \bar{X})\,\bar{X}] = 0$$

или

$$E(X_{r:n}\,\bar{X}) = E(\bar{X}^2) = \frac{1}{n}.$$

Сменом $n\bar{X} = \sum_{s=1}^{n} X_{s:n}$, добијамо да је $\sum_{s=1}^{n} \mu_{r,s:n} = 1$. □

Joshi и Balakrishnan (1981), су показали још нека занимљива својства матрице $(\mu_{r,s:n})$:

$$1.\ \sum_{s=r}^{n} \mu_{r,s:n} = 1 + \sum_{s=r}^{n} \mu_{r-1,s:n} \qquad r = 1, \dots, n$$

$$2.\ \sum_{s=r+1}^{n} \mu_{r,s:n} = \sum_{s=r+1}^{n} \mu_{s,s:n} - (n - r) \qquad r = 1, \dots, n-1$$

где је $\mu_{0,s:n} = 0$.

Доказ. Имамо

$$\sum_{s=r}^{n} \mu_{r,s:n} = 1 - \sum_{s=1}^{r-1} \mu_{r,s:n}$$

$$= 1 + \sum_{s=r}^{n} \mu_{r-1,s:n},$$

што можемо записати и на други начин

$$\mu_{r,r:n} = 1 + \sum_{s=r}^{n} \mu_{r-1,s:n} - \sum_{s=r}^{n} \mu_{r,s:n}.$$

На основу Тврђења 2.1. важи израз 1.

Сумирањем одговарућих израза за $\mu_{r+1,r+1:n}$ до $\mu_{n,n:n}$ добијамо израз 2. □

Пример 2.3. Одредићемо моменте статистика поретка са нормалном расподелом за $n = 2$.

$$\mu_{2:2} = E(X_{2:2}) = 2 \int_{-\infty}^{\infty} x\Phi(x)\phi(x)dx = -2 \int_{-\infty}^{\infty} \Phi(x)d\big(\phi(x)\big)$$

$$= 2 \int_{-\infty}^{\infty} \phi(x)^2 dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \exp(-x^2)dx = \frac{1}{\sqrt{\pi}}.$$

Како је $E(X_{1:2} + X_{2:2}) = E(X_1 + X_2) = 0$, добијамо

$$\mu_{1:2} = E(X_{1:2}) = -E(X_{2:2}) = -\frac{1}{\sqrt{\pi}}.$$

Пример 2.4. За $n = 5$ можемо да изразимо моменте статистика поретка коришћењем израза

$$I_n(a) = \int_{-\infty}^{\infty} [\Phi(ax)]^n e^{-x^2} dx,$$

тако да је $I_0(a) = \pi^{1/2}$.

Онда је

$$\int_{-\infty}^{\infty} \left[\Phi(ax) - \frac{1}{2}\right]^{2m+1} e^{-x^2} dx = 0 \qquad m = 0,1,2,\dots,$$

тако да је подинтегрална функција непарна по x.

Тако да важи

$$I_{2m+1}(a) = \frac{\sum_{i=1}^{2m+1}(-1)^{i+1}\binom{2m+1}{i} I_{2m-i+1}(a)}{2^i}.$$

Специјално,

$$I_1(a) = \frac{1}{2} I_0(a) = \frac{1}{2}\pi^{1/2},$$

$$I_3(a) = \frac{3}{2} I_2(a) - \frac{3}{4} I_1(a) + \frac{1}{8} I_0(a) = \frac{3}{2} I_2(a) - \frac{1}{4} I_0(a).$$

Диференцирањем интеграла $I_n(a)$ па a, за $n = 2$ добијамо следећу једнакост

$$(2\pi)^{1/2} I_2'(a) = \int_{-\infty}^{\infty} \Phi(ax) 2x e^{-\frac{x^2(a^2+2)}{2}} dx.$$

Парцијалном интеграцијом добијамо

$$I_2'(a) = \frac{1}{\pi^{1/2}} \frac{a}{(a^2+2)(a^2+1)^{1/2}}.$$

$$I_2(a) = \frac{1}{\pi^{1/2}} \arctan\left[(a^2+1)^{1/2}\right]$$

$$I_3(a) = \frac{3}{2\pi^{1/2}} \arctan\left[(a^2+1)^{1/2}\right] - \frac{1}{4}\pi^{1/2}.$$

Помоћу ових једнакости можемо израчунати моменте статистика поретка за $n = 5$ коришћењем парцијалне интеграције и једнакости $\phi'(x) = -x\phi(x)$ добијамо

$$E(X_{5:5}) = 5 \int_{-\infty}^{\infty} x\Phi^4(x)\phi(x)dx = 5 \int_{-\infty}^{\infty} 4\Phi^3(x)\phi(x)\phi(x)dx.$$

$$\mu_{5:5} = \frac{10}{\pi} I_3(1) = \frac{15}{\pi^{3/2}} \arctan\sqrt{2} - \frac{5}{2\pi^{1/2}} = \frac{5}{4\pi^{1/2}} + \frac{15}{2\pi^{1/2}} \arcsin\frac{1}{3} = 1{,}16296.$$

На сличан начин можемо добити и

$$\mu_{4:5} = \frac{5}{2\pi^{1/2}} - \frac{15}{\pi^{3/2}} \arcsin\frac{1}{3} = 0{,}49502$$

$$\mu_{3:5} = 0 \quad \mu_{2:5} = -\mu_{4:5} \quad \mu_{1:5} = -\mu_{5:5}.$$

Навешћемо резултате које је дао Nadarajah (2008) за $E(X_{r:n}^k)$, где случајан узорак $X_1, X_2, \ldots, X_n$ има нормалну расподелу. Као што је познато густина за $Y = X_{r:n}$ је

$$f_y(y) = \frac{n!}{(r-1)!\,(n-r)!}[\Phi(y)]^{r-1}[1-\Phi(y)]^{n-r}\phi(y).$$

k-ти моменат за $X_{r:n}^k$ се изражава на следећи начин

$$E(X_{r:n}^k) = \frac{n!}{(r-1)!\,(n-r)!}\int_{-\infty}^{\infty} y^k\,[\Phi(y)]^{r-1}[1-\Phi(y)]^{n-r}\phi(y)dy$$

$$= \frac{n!}{(r-1)!\,(n-r)!}\sum_{i=0}^{n-r}\binom{n-r}{i}(-1)^i\int_{-\infty}^{\infty} y^k\,[\Phi(y)]^{r+i-1}\phi(y)dy.$$

Коришћењем чињенице да је

$$\Phi(x) = \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right],$$

где је $\operatorname{erf}(x)$ функција грешке која је једнака

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt,$$

добијамо да је

$$E(X_{r:n}^k) = \frac{n!\,2^{1-r}}{\sqrt{2\pi}(r-1)!\,(n-r)!}\sum_{i=0}^{n-r}\binom{n-r}{i}\left(-\frac{1}{2}\right)^i\int_{-\infty}^{\infty} y^k\left[1+\operatorname{erf}\left(\frac{y}{\sqrt{2}}\right)\right]^{r+i-1}\exp\left(\frac{y^2}{2}\right)dy$$

$$= \frac{n!\,2^{1-r}}{\sqrt{2\pi}(r-1)!\,(n-r)!}\sum_{i=0}^{n-r}\binom{n-r}{i}\left(-\frac{1}{2}\right)^i\sum_{p=0}^{r+i-1}\binom{r+i-1}{p}$$

$$\cdot\int_{-\infty}^{\infty} y^k\left[\operatorname{erf}\left(\frac{y}{\sqrt{2}}\right)\right]^{r+i-1}\exp\left(\frac{y^2}{2}\right)dy$$

$$= \frac{n!\,2^{1-r}}{\sqrt{2\pi}(r-1)!\,(n-r)!}\sum_{i=0}^{n-r}\binom{n-r}{i}\left(-\frac{1}{2}\right)^i\sum_{p=0}^{r+i-1}\binom{r+i-1}{p}I(k,p).$$

Како је $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m x^{2m+1}}{(2m+1)m!}$,

Интеграл $I(k,p)$ се може изразити на следећи начин

$$I(k,p) = \int_{-\infty}^{\infty} y^k\left[\frac{2}{\sqrt{\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m x^{2m+1}}{(2m+1)m!}\right]^p\exp\left(\frac{y^2}{2}\right)dy$$

$$I(k,p)=\left(\frac{2}{\sqrt{\pi}}\right)^{p}\sum_{m_1=0}^{\infty}\dots\sum_{m_p=0}^{\infty}\frac{(-1)^{m_1+\cdots+m_p}}{2^{m_1+\cdots+m_p+\frac{p}{2}}(2m_1+1)\dots(2m_p+1)m_1!\dots m_p!}$$

$$\cdot\int_{-\infty}^{\infty}y^{2(m_1+\cdots+m_p)+p+k}\exp\left(\frac{y^2}{2}\right)dy.$$

$$\int_{-\infty}^{\infty}y^{2(m_1+\cdots+m_p)+p+k}\exp\left(\frac{y^2}{2}\right)dy=\begin{cases}2^{m_1+\cdots+m_p+\frac{p+k+1}{2}} & p+k \text{ парно}\\ 0 & p+k \text{ непарно}\end{cases}$$

па је интеграл $I(k,p)$ једнак

$$I(k,p)=\pi^{-\frac{p}{2}}2^{\frac{k+1}{2}}\sum_{m_1=0}^{\infty}\dots\sum_{m_p=0}^{\infty}\frac{(-1)^{m_1+\cdots+m_p}}{(m_1+1/2)\dots(2m_p+1/2)m_1!\dots m_p!}$$

$$\times\Gamma\left(m_1+\cdots+m_p+\frac{p+k+1}{2}\right),\text{за } p+k \text{ парно.}$$

Интеграл може да се поједностави коришћењем функције Lauricella (Exton, 1978),

$$F_A^{(n)}(a;b_1,\dots,b_n;c_1,\dots,c_n;x_1,\dots,x_n)$$
$$=\sum_{m_1=0}^{\infty}\dots\sum_{m_1=0}^{\infty}\frac{(a)_{m_1+\cdots+m_p}(b_1)_{m_1}\dots(b_n)_{m_n}x_1^{m_1}\dots x_n^{m_n}}{(c_1)_{m_1}\dots(c_n)_{m_n}m_1!\dots m_n!}$$

$$I(k,p)=\pi^{-\frac{p}{2}}2^{\frac{k+1}{2}+p}\Gamma\left(\frac{p+k+1}{2}\right)F_A^{(p)}\left(\frac{p+k+1}{2};\frac{1}{2},\dots,\frac{1}{2};\frac{3}{2},\dots,\frac{3}{2};-1,\dots,-1\right),$$
$$\text{за } k+p \text{ парно.}$$

Коначно добијамо израз

$$E(X_{r:n}^k)=\frac{n!\,2^{1-r}}{\sqrt{2\pi}(r-1)!\,(n-r)!}\sum_{i=0}^{n-r}\binom{n-r}{i}\left(-\frac{1}{2}\right)^{i}\sum_{p=0}^{r+i-1}\binom{r+i-1}{p}$$

$$\times\pi^{-\frac{p}{2}}2^{\frac{k+1}{2}+p}\Gamma\left(\frac{p+k+1}{2}\right)F_A^{(p)}\left(\frac{p+k+1}{2};\frac{1}{2},\dots,\frac{1}{2};\frac{3}{2},\dots,\frac{3}{2};-1,\dots,-1\right).$$

2.3. Моменти у случају дискретне расподеле

У случају дискретне расподеле k-ти редни моменат за $X_{r:n}$ може се директо добити из дефиниције

$$\mu_{r:n}^{(k)}=\sum_{x=0}^{\infty}x^k f_{(r)}(x),$$

где је $f_{(r)}(x)$ густина расподеле за $X_{(r)}$ коју смо већ дефинисали за дискретан случај (Feller,1957, p. 249). Нека је $q(x) = f(x+1) + f(x+2) + \cdots$, дефинишимо генераторне функције

$$P(s) = \sum_{x=0}^{\infty} f(x)s^x, \qquad Q(s) = \sum_{x=0}^{\infty} q(x)s^x.$$

За $|s| < 1$, k-ти извод од $P(s)$ је

$$P^{(k)}(s) = \sum_{x=k}^{\infty} x(x-1)\dots(x-k+1)f(x)s^{x-k}.$$

Ако k-ти факторијални моменат $\mu_{[k]}$ за X постоји, можемо да поставимо да је $s = 1$ и имамо

$$\mu_{[k]} = P^{(k)}(1).$$

За $|s| < 1$

$$Q(s) \cdot (1-s) = 1 - P(s),$$

одакле, диференцирањем k пута и коришћењем Лајбницове теореме добијамо

$$Q^{(k)}(s)(1-s) + kQ^{(k-1)}(s)(-1) = -P^{(k)}(s).$$

Ако $\mu_{[k]}$ постоји можемо да закључимо да важи $\mu_{[k]} = kQ^{(k-1)}(1)$. Специјално,

$$\mu_{[1]} = \mu = \sum_{x=0}^{\infty} q(x) = \sum_{x=0}^{\infty} [1 - F(x)],$$

$$\mu_{[2]} = E[X(X-1)] = 2\sum_{x=0}^{\infty} xq(x) = 2\sum_{x=0}^{\infty} x[1 - F(x)],$$

одакле следи да је $D(X) = \mu_{[2]} + \mu - \mu^2$.

Примењујемо ове рзултате на моменте за $X_{r:n}$

$$\mu_{r:n} = \sum_{x=0}^{\infty} \left[1 - I_{F(x)}(r, n-r+1)\right],$$

$$D(X_{r:n}) = 2\sum_{x=0}^{\infty} \left[1 - I_{F(x)}(r, n-r+1)\right] + \mu_{r:n} - \mu_{r:n}^2.$$

Специјално, добијамо за моменте екстремних вредности следеће

$$\mu_{n:n} = \sum_{x=0}^{\infty}[1 - F^n(x)], \qquad \mu_{1:n} = \sum_{x=0}^{\infty}[1 - F(x)]^n,$$

одатле добијамо

$$E(W_n) = \sum_{x=0}^{\infty}\{1 - F^n(x) - [1 - F(x)]^n\},$$

аналогно резултатима за непрекидну расподелу на интервалу $(0, \infty)$.

Такође важи

$$\mu_{r,s:n} = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} xy \int\int v^{r-1}(w-v)^{s-r-1}(1-w)^{n-s}dvdw.$$

Област по којој интегралимо је $\{(u,v): v \le w, F(x-1) \le v \le F(x), F(y-1) \le w \le F(y)\}$.

2.4. Рекурентне везе

Основне везе између момената које смо већ приказали моу да се користе као провера за директно израчунавање момената. Поред основних релација извешћемо и рекурентне везе момената.

Релација 1. За произвољну расподелу са коначним k −тим моментима важи

$$(n-r)\mu_{r:n}^{(k)} + r\mu_{r+1:n}^{(k)} = n\mu_{r:n-1}^{(k)}$$

где је $r = 1, \dots, n-1$ и $k = 1,2, \dots$

Основне резултате су добили Cole (1951) за непрекидан и Melnick (1964) за дискретан случај. Следећи доказ покрива оба случаја.

Доказ. (David, Joshi, 1968) Одабраћемо случајно један од $X_1, \dots, X_n$ и претпоставимо да је то $X_{i:n}$ $(i = 1, \dots, n)$. Резултујућа статистика поретка $X_{r:n-1}$ из случајног узорка обима $n-1$ је

$$X_{r:n-1} = X_{r:n+1} \quad \text{за } i = 1, \dots, r \quad \text{(A)}$$

$$X_{r:n-1} = X_{r:n} \quad \text{за } i = r+1, \dots, n \quad \text{(Б)}$$

како је (А) случајна величина са рангом $r+1$ у узорку обима n, а r у узорку $n-1$, итд. Одговарајуће вероватноће догађаја (А) и (Б) су r/n и $(n-r)/n$, тако да имамо следеће вероватноће

$$P\{ X_{r:n-1} \le x\} = P(\text{А})P\{ X_{r:n-1} \le x|\text{А}\} + P(\text{Б})P\{ X_{r:n-1} \le x|\text{Б}\}$$
$$= \frac{r}{n}P\{ X_{r+1:n} \le x\} + \frac{n-r}{n}P\{ X_{r:n} \le x\}$$

или

$$nF_{r:n-1}(x) = rF_{r+1:n}(x) + (n-r)F_{r:n}(x).$$

Коначност за μ_k је довољан не и неопходан услов, што ће важити и за остале релације.

За функцију $g(X)$, (cf. Srikantan, 1962), важи

$$(n-r)E[g(X_{r:n})] + rE[g(X_{r+1:n})] = nE[g(X_{r:n-1})].$$

Ово уопштење, обухвата функцију густине и генераторне функцкције момената, такође се односи и на релације 2 и 3, које су најважнији случајеви. □

Краће, али компликованије уопштење релације 1 добија се тако што $\mu_{r:n}^{(k)}$ запишемо као

$$\frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}\int_0^1 [F^{-1}(u)]^k u^{r-1}(1-u)^{n-r}du.$$

Последица 1А. За парно n важи

$$\frac{1}{2}\left(\mu_{\frac{1}{2}n+1:n}^{(k)} + \mu_{\frac{1}{2}n:n}^{(k)}\right) = \mu_{\frac{1}{2}n:n-1}^{(k)}.$$

Доказ. Заменимо у релацији 1 $r = \frac{1}{2}n$. □

За $k = 1$ очекиване вредности медијане су исте за узорак обима n (парно) и $n-1$.

Последица 1Б. Ако је расподела симетрична и n парно онда је

$$\mu_{\frac{1}{2}n:n-1}^{(k)} = \begin{cases} \mu_{\frac{1}{2}n:n}^{(k)}, & k \text{ парно} \\ 0, & k \text{ непарно.} \end{cases}$$

Доказ. У последицу 1А уведемо смену $\mu_{\frac{1}{2}n+1:n}^{(k)} = (-1)^k \mu_{\frac{1}{2}n:n}^{(k)}$,

Релација 2. За произвољну расподелу

$$\mu_{r:n}^{(k)} = \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1}\binom{n}{i}\mu_{i:i}^{(k)},$$

где је $r = 1, \dots, n$ и $k = 1,2, \dots$ Моменти за $X_{r:n}$ могу се изразити као моменти максимума из узорка од $r, \dots, n$. Ова релација се може утврдити применом релације 1.

Сада ћемо посматрати случај заједничке расподеле за $X_{r:n}$ и $X_{s:n}$. Избацивањем једне од првих r, затим једне од $s-r$ и на крају $n-s$, статистике поретка $X_{i:n}$ добијамо редом

$$X_{r:n-1} = X_{r+1:n} \quad X_{s:n-1} = X_{s+1:n} \quad \text{(В)}$$

$$X_{r:n-1} = X_{r:n} \quad X_{s:n-1} = X_{s+1:n} \quad \text{(Г)}$$

$$X_{r:n-1} = X_{r:n} \quad X_{s:n-1} = X_{s:n} \quad \text{(Д)}$$

Како су вероватноће за (В), (Г) и (Д) редом r/n, $(s-r)/n$, $(n-s)/n$ за било које x, y $(x \leq y)$ важи

$$nF_{r,s:n-1}(x,y) = rF_{r+1,s+1:n}(x,y) + (s-r)F_{r,s+1:n}(x,y) + (n-s)F_{r,s:n}(x,y).$$

Овај израз можемо да доведемо у везу са одговарајућим моментима производа и добијемо трећу релацију.

Релација 3. За било коју расподелу и $1 \leq r < s \leq n$

$$r\mu_{r1,s+1:n} + (s-r)\mu_{r,s1:n} + (n-s)\mu_{r,s:n} = n\mu_{r,s:n-1}.$$

Доказ. У формули

$$n\mu_{r-1,s-1:n-1} = \frac{n!}{(r-2)!\,(s-r-1)!\,(n-s)!} \int_{-\infty}^{\infty} \int_{-\infty}^{y} xy[F(x)]^{r-2} \times [F(y)-F(x)]^{s-r-1}[1-F(y)]^{n-s} dF(x)dF(y)$$

поделимо интеграл на суму три идентична интеграла, и добијамо

$$1 = F(x) + [F(y)-F(x)] + [1-F(y)].$$

Одатле непосредно следи релација 3.

3. Оцене индекса репа расподеле и високих квантила

Статистике поретка као што смо већ напоменули имају велики значај у статистичком закључивању. Користе се у оцењивању индекса репа расподеле и високих квантила. Индекс репа расподеле је индекс правилне промењивости који одређује дебљину репа расподеле од кога зависе многи математички модели у разним областима. Све оцене се дефинишу помоћу екстремних статистика поретка. У овом одељку ћемо приказати оцене индекса правилне промењивости α који се налази у функцијама расподела екстремних вредности (Фрешеова и Вејбулова). Правилна промењивост је локално својство функције, које одређује њено понашање у некој тачки.

Нека је $G_\gamma(x)$ функција расподеле екстремних вредности у γ-параметризацији

$$G_\gamma(x) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\},$$

где је $\gamma \in \mathbf{R}$ и $1 + \gamma x > 0$.

<u>Напомена.</u> γ-параметризација се добија увођењем смене $\gamma = 1/\alpha$ за Фрешеову и $\gamma = -1/\alpha$ за Вејбулову расподелу.

Нека је $(X_1, X_2, \dots, X_n)$ узорак из расподеле F која припада области привлачења функције $G_\gamma(x)$ за коју постоје константе $a_n > 0$ и $b_n \in \mathbf{R}$, такве да за сваки реалан број x важи

$$\lim_{n\to\infty} P\{M_n \le a_n x + b_n\} = G_\gamma(x)$$

где је $M_n = \max\{X_1, X_2, \dots, X_n\}$ и нека је $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ варијациони низ статисика поретка.

3.1. Пикандсова оцена

Пикандсова оцена параметра γ дефинише се на следећи начин, (Pickands, 1975)

$$\hat{\gamma}_n = \frac{1}{\ln 2} \ln \frac{X_{n-k+1:n} - X_{n-2k+1:n}}{X_{n-2k+1:n} - X_{n-4k+1:n}},$$

Наводимо резултате који су доказани у раду Dekkers, de Haan (1989).

Теорема 3.1. Нека је $(X_1, X_2, \dots, X_n)$ узорак из расподеле F и $\hat{\gamma}_n$ Пикандсова оцена. Ако важи $F \in D(G_\gamma)$ и $k \to \infty, \frac{k}{n} \to \infty$ кад $n \to \infty$ онда $\hat{\gamma}_n \to \gamma$ у вероватноћи кад $n \to \infty$.

Теорема 3.2. Нека је $(X_1, X_2, \dots, X_n)$ узорак из расподеле F и $\hat{\gamma}_n$ Пикандсова оцена. Ако важи $F \in D(G_\gamma)$ и $\frac{k}{n} \to \infty$ и $\frac{k}{\ln \ln n} \to \infty$ кад $n \to \infty$ онда $\hat{\gamma}_n \to \gamma$ скоро сигурно кад $n \to \infty$.

Теорема 3.3. Нека је $(X_1, X_2, \dots, X_n)$ узорак из расподеле F и $\hat{\gamma}_n$ Пикандсова оцена. Ако важи $\in D(G_\gamma)$. Израз

$$U(x) = \left\{\frac{1}{1 - F(x)}\right\}^{-1}$$

је уопштени инверз и претпоставимо да функција $U(x)$ има позитиван извод и да постоји позитивна функција $a(t), t \in \mathbf{R}$ таква да за свак позитиван број $\mathbf{R}$ важи једнакост

$$\lim_{t \to \infty} \frac{(tx)^{1-\gamma} U'(tx) - t^{1-\gamma} U'(t)}{a(t)} = \ln x$$

или

$$\lim_{t \to \infty} \frac{(tx)^{1-\gamma} U'(tx) - t^{1-\gamma} U'(t)}{a(t)} = -\ln x$$

тада случајна величина $\sqrt{m}(\hat{\gamma}_n - \gamma)$ има асимптотску нормалну расподелу $N(0, \sigma^2)$, (Dekkers, de Haan, 1989), где је

$$\sigma^2 = \frac{\gamma^2 (2^{2\gamma+1} + 1)}{\{2(2^\gamma - 1)\ln 2\}^2}.$$

Оцена квантила ξ_1 дефинише се на следћи начин

$$\hat{\xi}_1 = \frac{X_{n-k+1:n} - X_{n-2k+1:n}}{\left(\frac{1}{2}\right)^{\hat{\gamma}_n} - 1} + X_{n-k+1:n},$$

и важи

$$\sqrt{2k}\, \frac{\hat{\xi}_1 - \xi_1}{X_{n-k+1:n} - X_{n-2k+1:n}} \xrightarrow{d} N\left(0, \frac{3\gamma^2 2^{2\gamma-1}}{(2^\gamma - 1)^6}\right).$$

За мало p близу 0, оцена за ξ_{1-p} је

$$\hat{\xi}_{1-p} = \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}_n} - 1}{1 - \left(\frac{1}{2}\right)^{\hat{\gamma}_n}} (X_{n-k+1:n} - X_{n-2k+1:n}) + X_{n-k+1:n},$$

За $k = [np]$ важи, (Dekkers, de Haan, 1989),

$$\sqrt{2k}\, \frac{X_{n-k+1:n} - \xi_{1-p}}{X_{n-k+1:n} - X_{n-2k+1:n}} \xrightarrow{d} N\left(0, \frac{\gamma^2 2^{2\gamma+1}}{(2^\gamma - 1)^2}\right).$$

3.2. Хилова оцена

Нека је $(X_1, X_2, \ldots, X_n)$ узорак из расподеле F која припада области привлачења функције расподеле екстремних вредности G_γ при чему је параметар $\gamma = 1/\alpha$ позитиван. У Хиловој оцени фигурише k максималних статистика поретка. Хилова ооцена параметра γ се дефинише на следећи начин, (Hill, 1975),

$$\hat{H}_n = \frac{1}{k}\sum_{i=1}^{k} \ln X_{n-i+1:n} - \ln X_{n-k:n}.$$

Теорема 3.4. (Mason, 1982) Оцена $\hat{H}_n$ је слабо конзистентна оцена параметра γ ако важе следећи услови $\to \infty$ и $\frac{k}{n} \to \infty$ кад $n \to \infty$.

Теорема 3.5. (Deheuvels, Haeusler, Mason, 1988) Оцена $\hat{H}_n$ је јако конзистентна оцена параметра γ ако важе следећи услови $\frac{k}{n} \to \infty$ и $\frac{k}{\ln \ln n} \to \infty$ кад $n \to \infty$.

Под одређеним условима случајна величина $\sqrt{m}(\hat{H}_n - \gamma)$ кад $n \to \infty$ има нормалну расподелу $N(0, \sigma^2)$. Резултат о асимптотској нормалности Хилове оцене омогућава одређивање интервала поверења непознатог параметра γ.

Закључак

Описана су елементарна својства статистика поретка. Представљен је један део ове широке теме. Обухваћеме су расподеле статистика поретка из непрекидне и дискретне популације, њихови моменти и најчешће коришћене оцене индекса репа расподеле. Рад се бави статистикама поретка које су независне и равномерно распоређене, што дата дефиниција не захтева. Међутим, веома велики број књига и научних радова који се баве овом темом је управо фокусиран на такве статистике поретка. Важно је напоменути да иако постоје сличности између статистика поретка из непрекидне и дискретне популације, постоје својства која задовољавају само статистике поретка из непрекидне расподеле, док не важе за оне из дискретне расподеле. На пример, видели смо да, у непрекидном случају, статистике поретка формирају ланац Маркова, док у дискретном случају то не важи увек, већ само под одређеним условима. Као што је већ речено, примена статистика поретка је веома велика. Најчешће се користе у оцењивању параметара специфичних расподела, у теорији расподела екстремних вредности, у оцењивању коефицијената регресије, као и за откривање аутлајера. Корисне су не само за оцењивање већ и за предвиђање и одређивање интервала поверења. Линеарне функције статистика поретка, као што је распон, и екстреми су применљиви у разним областима, на пример у хидрологији, метеорологији, сеизмологији, економетрији и многим другим областима.

Литература

Дэйвид, Г. (1979): *Порядаковые статистики*. Наука, Москва.
David, H.A., Nagaraja H.N. (2003): *Order statistics*. Wiley, New Jersey.
Balakrishnan, N., Rao C.R. (1998): *Order statistics,Theory & Methods*. Elsevier.
Младеновић, Павле (2002): *Екстремне вредности случајних низова*. Математички факултет, Београд.
Јевремовић, Весна (2009): *Вероватноћа и статистика*. Математички факултет, Београд.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. Wiley, New York
Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998). *Records*, Wiley, New York, [13]
Chu, J. T. (1957). Some uses of quasi-ranges. *Ann. Math. Statist.* **28**, 173-80. [15]
Cole, R. H. (1951). Relations between moments of order statistics. . *Ann. Math. Statist*. **22**, 308-10, [30]
David, F. N and Johnson, N. L. (1954). Statistical treatment of censored data. Fundamental formulae. *Biometrika* **42**, 228-40. [19]
David, H. A. and Joshi, P. C. (1968). Recurrence relations between moments of order statistics for exchangeable variates. *Ann. Math. Statist* **39**, 272-4 [30]
Dekkers, A. L. M. and de Haan, L. (1989). On the estimation of the extremvalue index and large quantile estimation. *Ann. Statist*. 17, 1795-55. [33,34]
Feller , W. (1957). *An Introduction to Probability Theory and Its Aplications,* Vol. I, 2nd ed. Wiley, New York. [29]
Hill, B. M. (1975). A simple general approach to inference about tail of a distribution. *Ann. Statist*. **3**, 1163-74. [35]
Jones, M. C. (2002). On fractional uniform order statistics. *Statist. Probab. Lett.***58**, 93-6 [13]
Joshi, P. C. and Balakrishnan, N. (1981). An identity for the moments of normal order statistics with aplications. *Scand. Actuarial J. 1981*. 203-13. [23]
Joshi, P. C. and Balakrishnan, N. (1982). Recurrence relations and identities for the product moments of order statistics. *Sankhya B.* **44,** 39-49. [23,24]
Khatri, C/ G. (1962). Distribution of order statistics for discrete case. *Ann. Inst. Statist. Math.* **14**, 167-71. [6]
Malmquist, S. (1950). On a property of order statistics from a rectangular distribution. *Skand. Aktuarietidskr.* **33**, 214-22. [12]
Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab*. **10**, 754-64. [34]
Melnik, E. L. (1964). Moments of ranked Poisson variates. M. S. thesis, irginia Polytechnic Institute. [30]
Nagaraja, H. N. (1982). On the non-Marcovian structure of discrete order statistics. *J. Statist. Plann. Inf.* **7**, 29-33. [12]
Nadarajah, S. (2008). Explicit expressions for moments of order statistics, *Statist. Probab. Lett.* **28,** 196-205, [26]
Pickands, J. , III (1975). Statistical inference using extreme-order statistics. *Ann. Statist*. **3**, 119-31. [33]
Pinsker, I. Sh., V. Kipnisand E. Grechanovsky (1986). Arecursive formula for the probability occurrence of at least m out of Nevents. *Amer. Statist.* **40**, 275-276. [5]
Renyi, A.(1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung*. **4**,191-231. [11]
Scheffe, H. and Tukey, J. W. (1945). Non-parametric estimation. Validation of order statistics. Ann. Math. Statist. 16, 187-92. [6,15]
Sen, P. K. (1959). On the moments of the sample quantiles. *Calcutta Statist. Ass. Bull.* **9**, 1-19. [18]
Serfling, R. J. (1980). Approximation Theorems od Mathemaical Statistits. Wiley, New York. [7]
Srikantan, K. S. (1962). Recurrence relations between the PDF's of order statistics, and some aplications. *Ann. Math.Statist.* **33**, 169-77. [31]

Stigler, S. M. (1977). Fractional order statistics, with aplicatins. *J. Amer. Statist. Ass.* **72,** 544-50 [13]
Sukhatme, P. V. (1937). Tests of significance for samples of χ^2 population with two dwgrees of freedom. *Ann. Eugen.* **8**, 52-6. [10,11]
Thompson, W. R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown form. *Ann. Math. Statist*. **7,** 122-8, [15]
Tippet, L. H.C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika* **17**, 364-87 [21,22]
Wilks, S. S. (1962). *Mathematical Statistics*. Wiley, New York [16]