УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Никола Митриновић

Неједнакости за U-статистике с недегенерисаним језгром

— мастер–рад —

Београд, 2017.

Садржај

1 Основни појмови и примери **1**

2 Мартингали и U-статистике **6**

3 Неједнакости за U-статистике **22**

3.1 Неједнакости са моментима 22

3.2 Максималне неједнакости 26

4 Примена неједнакости на доказ централне граничне теореме **29**

Поглавље 1

Основни појмови и примери

Нека је $\mathcal{P}$ фамилија вероватносних мера на произвољном мерљивом простору $(\Omega,\mathcal{A})$. Проблеми којима ћемо се бавити су непараметарски, што значи да ће $\mathcal{P}$ бити велика фамилија расподела која подлеже само слабим условима као што су непрекидност и постојање момената. Нека је $\Omega(P)$ реална функција дефинисана за $P \in \mathcal{P}$. Први појам који ћемо дефинисати је оцењиви параметар (регуларни параметар).

Дефиниција 1.0.1 *Кажемо да је функција $\theta(P)$ оцењиви параметар на $\mathcal{P}$, ако постоји природан број m и непристрасна оцена за $\theta(P)$ која садржи m независних једнакорасподељених случајних величина које припадају P, тачније, ако постоји реална мерљива функција $h(x_1, x_2, \ldots, x_m)$ таква да важи $E_P(h(X_1, X_2, \ldots, X_m)) = \theta(P)$, за све $P \in \mathcal{P}$, где су $X_1, X_2, \ldots, X_m$ независне једнакорасподељене случајне величине са расподелом P. Најмањи природан број m са овим својством назива се ред оцене $\theta(P)$.*

За функцију h увек можемо претпоставити да је симетрична, јер ако је f непристрасна оцена за $\theta(P)$, тада је аритметичка средина функције f по свим пермутацијама такође непристрасна, али се добија и њена симетричност. Дакле:

$$h(x_1, x_2, \ldots, x_m) = \frac{1}{m!} \sum_{\pi \in \Pi_m} f(x_{\pi_1}, \ldots, x_{\pi_m}).$$

Сумирање се врши по свим пермутацијама скупа од m елемената. Функције h је очигледно симетрична по својим аргументима и има исто очекивање као f у смислу расподеле P.

Дефиниција 1.0.2 *За реалну мерљиву функцију $h(x_1, x_2, \ldots, x_m)$ и узорак $X_1, X_2, \ldots, X_n$ димензије $n(n \geqslant m)$ из расподеле P, U-статистика са језгром h дефинисана је следећим изразом*

$$U_n = U_n(h) = \frac{(n-m)!}{n!} \sum_{P_{m,n}} h(X_{i_1}, \ldots, X_{i_m}).$$

Дакле, из узорка димензије n бирамо m чланова, а затим вршимо пермутације тог новоизабраног скупа. Сумирање иде по свим тим скуповима којих има

$$\binom{n}{m} m! = \frac{n!}{(n-m)!m!} m! = \frac{n!}{(n-m)!}.$$

Претходни случај је када језгро h није симетрично. Када јесте симетрично тада статистика U_n добија облик

$$U_n = U_n(h) = \frac{1}{\binom{n}{m}} \sum_{C_{m,n}} h(X_{i_1}, \ldots, X_{i_m}),$$

јер овај пут не морамо да вршимо пермутације изабраних m-торки због симетричности, па се симулација врши само по свим могућим m-торкама из узорка. Ако $\theta(P) = E_P(h(X_1, X_2, \ldots, X_m))$ постоји за све $P \in \mathcal{P}$ онда је очигледно својство U-статистике U_n, да је непристрасна оцена за $\theta(P)$, што ћемо касније и доказати.

Пример 1 *Моменти (преузет из [3])*

Ако $\mathcal{P}$ обухвата све расподеле које имају коначно математичко очекивање, тада је очекивање $\mu = \mu(P) = \int x dP(x)$ оцењиви параметар реда $m = 1$, јер је $f(X_1) = X_1$ непристрасна оцена за μ. Одговарајућа U-статистика је узорачка статистика, $U_n = \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i^n$. Слично ако $\mathcal{P}$ обухвата све расподеле које узимају реалне вредности и имају коначне k-те моменте, $\mu_k = \int x^k dP(x)$ је оцењиви параметар степена 1, чија је U-статистика узорачки моменат k-тог реда, $\frac{1}{n} \sum_{i=1}^n X_i^k$

Ако желимо да оценимо квадрат очекивања односно $\theta(P) = \mu^2$, да бисмо дошли до фунцкије h, приметимо да је $E(X_1 X_2) = \mu^2$. На основу тога можемо направити U-статистику са језгром $h(x_1, x_2) = x_1 x_2$ на следећи начин:

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j} X_i X_j = \frac{2}{n(n-1)} \sum_{i > j} X_i X_j$$

Приметимо да је језгро реда 2 и може се доказати да не може бити реда 1.

Ако узмемо $\mathcal{P}$ такво да обухвата све расподеле које имају коначне моменте другог реда, онда је дисперзија $\sigma^2 = \mu_2 - \mu^2$ такође оцењив параметар реда 2, јер μ_2 можемо оценити са X_1^2, а μ^2 са X_1X_2. Међутим језгро није симетрично ($f(x_1, x_2) = x_1^2 - x_1x_2$), али се једноставно симетризује

$$h(x_1, x_2) = \frac{1}{2}(f(x_1, x_2) + f(x_2, x_1)) = \frac{(x_1 - x_2)^2}{2}.$$

Када смо нашли симетрично језгро, лако је направити U-статистику

$$U_n = \frac{2}{n(n-1)} \sum_{i<j} \frac{(X_i - X_j)^2}{2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

ово је заправо поправљена узорачка дисперзија.

Као закључак треба приметити да је и линеарна комбинација оцењивих параметара поново оцењив параметар, као што смо и видели у овом примеру.

Пример 2 *Емпиријска фунцкија расподеле (преузет из [2])*
Нека је тражени параметар $F(t), t \in \mathbb{R}$, произвољне функције расподеле F. Приметимо да је то параметар реда 1, јер га можемо оценити са $f(X_1) = I(X_1 < t)$. То је симетрично језгро, а одговарајућа U-статистика

$$U_n = \frac{1}{n} \sum_{i=1}^{n} I(X_i < t),$$

што је заправо одговарајућа емпиријска функција расподеле.

Пример 3 *Вилкоксонов тест означених рангова*
Нека је F непрекидна расподела симетрична око неког $\theta \in \mathbb{R}$. Разматрамо проблем тестирања хипотезе H_0, да је расподела F симетрична у односу на θ_0, на основу узорка $X_1, \ldots, X_n$ из F. Одговарајући тест је Вилкоксонов тест означених рангова. Нека је $Z_i = X_i - \theta_0$, а тест статистика је онда

$$W_n^+ = \frac{1}{2} \sum_{i=1}^{n} R_i^+ (1 + \operatorname{sgn}(Z_i)),$$

где је R_i^+ ранг, односно редни број $|Z_i|$ међу $|Z_1|, \ldots, |Z_n|$. Иако W_n^+ није U-статистика, може се представити као линеарна комбинација две U-статистике на следећи начин

$$W_n^+ = \sum_i I(Z_i > 0) + \sum_{i<j} I(Z_i + Z_j > 0),$$

и, када се запише у том облику, можемо боље испитати њено понашање. Прва U*-статистика има језгро* $h(z) = I(z > 0)$*, док цела* U*-статистика има облик* $U_n^{(1)} = \frac{1}{n}\sum_1^n I(Z_i > 0)$*. Ово је статистика која се користи за обични тест знакова. Друга* U*-статистика има језгро* $h(x_1, x_2) = I(z_1 + z_2 > 0)$ *док сама статистика има облик* $U_n^{(2)} = \binom{n}{2}^{-1}\sum_{i<j} I(Z_i + Z_j > 0)$*. Напокон долазимо до*

$$W_n^+ = nU_n^{(1)} + \binom{n}{2}U_n^{(2)}.$$

Из ове репрезентације и под претпоставком нулте хипотезе добија се очекивање

$$EW_n^+ = nP(X_1 > \theta_0) + \binom{n}{2}P(X_1 + X_2 > 2\theta_0) = \frac{n(n+1)}{4}.$$

За узорке малог обима постоје таблице расподеле ове статистике. За узорке обима већег од 20, расподела се може апроксимирати нормалном расподелом, што ћемо објаснити у даљем излагању.

Пример 4 *Тестирање симетрије*

У неким ситуацијама, битно је тестирати симетрију без знања центра. Сада ћемо приказати један метод базиран на узорку обима 3, X_1, X_2, X_3*, из непрекидне расподеле, симетричне око тачке* ξ*. Тада је*

$$P(X_1 > \frac{X_2 + X_3}{2}) = P(X_1 - \xi > \frac{(X_2 - \xi) + (X_3 - \xi)}{2}) = \frac{1}{2}.$$

Због овога је $f(X_1, X_2, X_3) = \operatorname{sgn}(2X_1 - X_2 - X_3)$ *непристрасна оцена за* $\theta(P) = P(2X_1 > X_2 + X_3) - P(2X_1 < X_2 + X_3)$*. Када је* P *симетрична расподела, тада* $\theta(P)$ *има вредност 0. Ако бисмо желели да направимо симетрично језгро то можемо урадити овако:*

$$h(x_1, x_2, x_3) = \frac{1}{3}(f(x_1, x_2, x_3) + f(x_2, x_3, x_1) + f(x_3, x_1, x_2)).$$

Ово је пример језгра реда 3. Хипотеза симетрије се одбацује за велику апсолутну вредност одговарајуће U*-статистике са језгром* h*. Такође* h *можемо записати и у следећем облику:*

$$h(x_1, x_2, x_3) = \frac{1}{3}(\text{med}(x_1, x_2, x_3) - \frac{x_1 + x_2 + x_3}{3}),$$

где функција med означава медијану. Дакле валидност теста такође следи из тога што је медијана узорка димензије 3 са истом вероватноћом већа, односно мања од аритметичке средине тог узорка.

Пример 5 *Тестирање експоненцијалности (преузети из [1])*

Нека су $X_1, \ldots, X_n$ позитивне, независне, једнакорасподељене случајне величине са функцијом расподеле F. Можемо тестирати да ли је F експоненцијална функција расподеле, помоћу параметра

$$\theta = \int_0^\infty \int_0^\infty (1 - F(s+t))dF(s)dF(t).$$

Познато је да је F експоненцијална функција расподеле ако и само ако задовољава услов

$$1 - F(s+t) = (1 - F(s))(1 - F(t)).$$

Дакле, под претпоставком H_0, важи

$$\theta = (\int_0^\infty (1 - F(t))dF(t))^2 = (\frac{1}{2}\int_0^\infty 2\lambda e^{-2\lambda t} dt)^2 = \frac{1}{4}.$$

Дефинишемо $f(x, y, z) = I(x - y - z > 0)$ тако да је

$$\theta = \int_0^\infty \int_0^\infty \int_0^\infty f(x, y, z)dF(x)dF(y)dF(z),$$

Лако се проверава да је овај троструки интеграл заиста једнак $\frac{1}{4}$ то јест нашем параметру. За крај овог примера направимо симетрично језгро реда 3

$$h(x, y, z) = \frac{1}{3}(f(x, y, z), f(y, z, x), f(z, x, y)).$$

Сада смо добили симетрично језгро, а самим тим и одговарајућу U-статистику коју можемо користити за тестирање експоненцијалности.

Поглавље 2

Мартингали и U-статистике

За мартингале, специјалне врсте случајних процеса постоје граничне теореме које говоре о њиховој конвергенцији. Овим проблемом се бавио амерички математичар Дуб и дошао до закључка да под неким условима постоје граничне вредности у L_1 смислу и скоро сигурно и за супермартингале и за обрнуте супер мартингале. Биће доказано да је свака U-статистика обрнути мартингал, па ће граничне теореме за мартингале бити од користи при испитивању граничне вредности U-статистика. Пре свега морамо дефинисати поменуте појмове.

Дефиниција 2.0.1 *Низ случајних величина $\{X_n, n \geqslant 1\}$ је мартингал у дискретном времену у односу на растући низ σ-алгебри $\{\mathcal{F}_n, n \geqslant 1\}$, ако су испуњени следећи услови:*
а)$E|X_n| < +\infty$, за свако $n \geqslant 1$;
б)X_n је $\mathcal{F}_n$-мерљива за свако $n \geqslant 1$;
в)$E(X_m|\mathcal{F}_n) = X_n$, за свако $m \geqslant n \geqslant 1$.

Аналогно добијамо дефиницију супермартингала, ако уместо услова *в*) важи
в') $E(X_m|\mathcal{F}_n) \leqslant X_n$, за свако $m \geqslant n \geqslant 1$,
односно субмартингала ако важи
в") $E(X_m|\mathcal{F}_n) \geqslant X_n$, за свако $m \geqslant n \geqslant 1$.
Из дефиниције следи да је сваки мартингал истовремено субмартингал и супермартингал.

Дефиниција 2.0.2 *Низ случајних величина $\{X_n, n \geqslant 1\}$ је обрнути мартингал у дискретном времену у односу на опадајући низ σ-алгебри $\{\mathcal{F}_n, n \geqslant 1\}$, ако су испуњени следећи услови:*
а)$E|X_n| < +\infty$, за свако $n \geqslant 1$;
б)X_n је $\mathcal{F}_n$-мерљива за свако $n \geqslant 1$;

в)$E(X_m|\mathcal{F}_n) = X_n$, *за свако* $n \geqslant m \geqslant 1$.

За низ случајних величина $X_1, X_2, \dots$ кажемо да задовољава услов симетрије ако су коначнодимензионе расподеле инваријантне при пермутацији индекса елемената. Тачније за свако $n \geqslant 1$ и сваку пермутацију $\pi \in \Pi$, важи

$$(X_{\pi_{(1)}}, \dots, X_{\pi_{(n)}}) \stackrel{d}{=} (X_1, \dots, X_n).$$

Уведимо појам σ-алгебра пермутабилних догађаја. Нека је дат низ случајних величина $X = (X_1, X_2, \dots)$ који задовољава услов симетрије. Можемо га посматрати као функцију са доменом Ω и кодоменом $\mathbb{R}^{\mathbb{N}}$. Уколико нас занимају само вредности које узимају случајне величине $(X_1, X_2, \dots)$, σ-алгебра $\mathcal{A}$ над Ω може бити преширока, пошто се неким различитим $\omega \in \Omega$ функцијом X додељују исте вредности. Довољно је посматрати σ-алгебру генерисану скуповима $X^{-1}(B)$, где је B елемент Борелове σ-алгебре над $\mathbb{R}^\infty$, $\mathcal{B}^\infty$. Могу нас занимати и само симетрични скупови у $\mathbb{R}^\infty$, то јест они реални низови код којих се произвољном пермутацијом коначно много елемената добија исти тај низ. За такве скупове $C \subseteq \mathcal{B}^\infty$, који такође чине σ-алгебру над $\mathbb{R}^\infty$, одговарајућа σ-алгебра над Ω је дефинисана са $X^{-1}(C)$.

Обележимо са Π_n скуп пермутација које мењају само неке од првих n елемената. Тачније:

$$\Pi_n = \{\pi : \mathbb{N} \to \mathbb{N} : \pi(i) = i, i > n\}.$$

Дефинишемо σ-алгебре на $\mathbb{R}^\infty$ са

$$\varepsilon_n = \{A \in \mathcal{B}^\infty : \pi^{-1}(A) = A, \forall \pi \in \Pi_n\}.$$

Како је $\Pi_n \subset \Pi_{n+1}$, то је и $\varepsilon_{n+1} \subset \varepsilon_n$, дакле имамо опадајући низ σ-алгебри и има смисла дефинисати $\cap_{n=1}^{\infty} \varepsilon_n$ - σ-алгебру пермутабилних догађаја.

Теорема 2.1 *(Хјуит-Савиџ закон 0-1)*
Ако за низ случајних величина важи услов симетрије, онда је одговарајућа σ-алгебра пермутабилних догађаја тривијална.

Приметимо да је Колмогоровљев закон 0-1, који тврди да је σ-алгебра генерисана репом низа случајних величина тривијална, специјални случај претходне теореме, јер је она подскуп σ-алгебре свих пермутабилних догађаја.

Теорема 2.2 *Нека је низ случајних величина $\{X_n, n \geqslant 1\}$ обрнути мартингал за који важи $E(|X_n|) < +\infty$, тада је*

$$\lim_{n\to\infty} X_n = E(X_0| \cap_{n=1}^{+\infty} \mathcal{F}_n)$$

скоро сигурно и у L_1 смислу.

Докази претходне две теореме се налазе редом у [4], [1].

Ради једноставности записа уведимо нову ознаку за вектор статистика поретка $X^{(n)} = (X_{(1)}, \dots, X_{(n)})$.

Лема 2.0.1 *Нека је $U_n(h), (n \geqslant m)$ низ U-статистика базираних на низу независних једнакорасподељених случајних величина $\{X_k, k \geqslant 1\}$ и симетричном језгру h. Нека је дефинисана филтрација $\{\mathcal{F}_n, n \geqslant 1\}$ тако да је $\mathcal{F}_n = \sigma\{X^{(n)}, X_{n+1}, X_{n+2}, \dots\}$. Тада за $n \geqslant m$ важи*

а)$\mathcal{F}_{n+1} \subset \mathcal{F}_n$;

б)$U_n(h)$ је $\mathcal{F}_n$-мерљива функција;

в)$U_n(h) = E(U_l|\mathcal{F}_n)$, за $m \leqslant l \leqslant n$.

Доказ

а) Ово тврђење можемо интуитивно разумети на следећи начин: довољно је посматрати инверзне слике скупова из $\mathcal{B}^{n+1}$. Инверзна слика једног таквог скупа у односу на случајне величине $(X_{(1)}, \dots, X_{(n+1)})$ је подскуп инверзне слике истог скупа у односу на $(X_{(1)}, \dots, X_{(n)}, X_{n+1})$, јер је у првом случају постављен јачи услов због додатне статистике поретка.

б) Пошто је h симетрично језгро можемо га посматрати као функцију првих n статистика поретка, а како је h мерљива функција, $U_n(h)$ је мерљива у односу на σ-алгебру $\sigma\{X^{(n)}\}$, па и у односу на $\mathcal{F}_n$.

в) Пошто су X_i независне и једнакорасподељене важи

$$E(h(X_{i_1}, \dots, X_{i_m})|\mathcal{F}_n) = E(h(X_1, \dots, X_m)|\mathcal{F}_n).$$

За сваки подскуп $\{i_1, \dots, i_m\}$ скупа $\{1, 2, \dots, n\}$. Даље примењујући мерљивост функције $U_n(h)$, дефиницију U-статистике и добијену последицу симетрије, добијамо:

$$U_n(h) = E\left(U_n(h)|\mathcal{F}_n\right) = E\left(\binom{n}{m}^{-1} \sum_{1\leqslant i_1<\dots<i_m\leqslant n} h(X_{i_1}, \dots, X_{i_m})|\mathcal{F}_n\right)$$

$$= E(h(X_1, \dots, X_m)|\mathcal{F}_n).$$

Најзад, за неко $l, m \leqslant l \leqslant n$, примењујући телескопско својство условног очекивања имамо:

$$E(U_l|\mathcal{F}_n) = E(E(h(X_1,\dots,X_m)|\mathcal{F}_l)|\mathcal{F}_n) = E(h(X_1,\dots,X_m)|\mathcal{F}_n) = U_n(h).$$

$\square$

Резултат претходне леме је да је низ U-статистика, $\{U_n(h), n \geqslant m\}$ обрнути мартингал у односу на дату филтрацију.

Теорема 2.3 *Нека је $U_n(h), n \geqslant m$ низ U-статистика. Тада важи*

$$U_n \to \theta(P), n \to +\infty,$$

где је конвергенција скоро сигурно и у L_1 смислу.

Доказ

Према леми 2.0.1, овај низ U-статистика је обрнути мартингал, па користећи теорему о конвергенцији обрнутих мартингала, имамо

$$\lim_{n\to+\infty} U_n(h) = E(U_m(h)| \cap_{n=m}^{+\infty} \mathcal{F}_n),$$

скоро сигурно и у L_1 смислу. На основу Хјуит-Савиџ закона 0-1, пошто је σ-алгебра $\cap_{n=m}^{+\infty}\mathcal{F}_n$ инваријантна у односу на пермутације, она је тривијална, а условно очекивање у односу на такву σ-алгебру је једнако обичном очекивању. Тиме добијамо тражено:

$$\lim_{n\to+\infty} U_n(h) = E(U_m(h)) = \theta.$$

Потребно је још образложити зашто је $\cap_{n=m}^{+\infty}\mathcal{F}_n$ инваријантна у односу на пермутације. Нека је $A \in \mathcal{F}_n$. Тада је он облика

$$A = (X_{(1)},\dots,X_{(n)},X_{n+1},\dots)^{-1}(B),$$

за неко $B \in \mathcal{B}^\infty$. Наравно, $A = (X_1, X_2, \dots)^{-1}(B_1)$ за неко $B_1 \supset B$, пошто немамо ограничења по координатама. Такође, пошто без статистика поретка није битан распоред координата

$$A = (X_{j_1},\dots,X_{j_n},X_{n+1},\dots)^{-1}(B_1).$$

Дакле, $A \in \varepsilon_n$. Одавде смо добили да је σ-алгебра $\cap_{n=m}^{+\infty}\mathcal{F}_n$ подскуп σ-алгебра свих пермутабилних скупова.

Како скоро сигурна конвергенција имплицира конвергенцију у вероватноћи, резултат ове теореме је да је U-статистика постојана оцена за дати параметар. Још један резултат који описује U-статистике као оцене параметара даје следећа теорема

Теорема 2.4 *За произвољну непристрасну оцену S параметра $\theta(P)$ постоји U-статистика U таква да важи*

$$DU \leqslant DS,$$

а једнакост важи ако је $U = S$ скоро сигурно. Додатно, ако је фамилија расподела $\mathcal{P}$ и комплетна, онда је U-статистика

$$U = \frac{1}{n!}\sum_{\pi\in\Pi} S(X_{\pi(1)}, \ldots, X_{\pi(n)}),$$

јединствена најбоља оцена параметра $\theta(P)$.

Доказ Пошто је вектор статистика поретка довољна статистика за сваку фамилију расподела $\mathcal{P}$ други део теореме је последица познате теореме Леман-Шефе о јединственој најбољој оцени. За доказ првог дела теореме приметимо да је статистика U, дефинисана у исказу, једнака условном очекивању $E(S|X^{(n)})$. Заиста, ако знамо вредности статистика поретка, због независности елемената у узорку, сваки распоред вредности из варијационог низа, по случајним величинама $X_1, \ldots, X_n$ је једнаковероватан, са вероватноћом $(n!)^{-1}$, а тиме је одређена и вредност статистике S. Применом Јенсенове неједнакости имамо

$$EU^2 = E((E(S|X^{(n)}))^2) \leqslant E(E(S^2|X^{(n)})) = ES^2.$$

Једнакост важи ако и само ако је S мерљива у односу на $\sigma\{X^{(n)}\}$, односно ако је $U = S$ скоро сигурно, а одавде због непристрасности оцена U и S, следи неједнакост дисперзија.

$\square$

Можемо приметити да је први део теореме специјални случај теореме Рао-Блеквел о побољшању оцене коришћењем условног очекивања.

Дефиниција 2.0.3 *Ако имамо симетрично језгро h реда m које је оцена за параметар $\theta(P)$, кажемо да је језгро h дегенерисано у мери P, ако за неко $m_0 < m$ и све $x_{m_0+1}, \ldots, x_m$ важи*

$$\int \cdots \int h(x_1, \ldots, x_m) dP(x_1) \ldots dP(x_{m_0}) = 0,$$

иначе кажемо да је језрго h недегенерисано. Појам дегенерисаности у даљем тексту ће се увек односити на случај $m_0 = 1$, дакле када је интеграл по једном, било ком, аргументу једнак нули.

Постматрајмо симетрично језгро h реда m које је интеграбилно у односу на производ мера, то јест

$$\int \cdots \int |h(x_1, \ldots, x_m)| dP(x_1) \ldots dP(x_m) < +\infty.$$

Дефиниција 2.0.4 *Дефинишимо, за $0 \leqslant c \leqslant m$ функције*

$$\tilde{h}_c(x_1, \ldots, x_c) = Eh(x_1, \ldots, x_c, X_{c+1}, \ldots, X_m) =$$

$$= \int \cdots \int h(x_1, \ldots, x_m) dPx_{c+1} \ldots dPx_m.$$

Приметимо да за тако дефинисане функције важе следеће особине:
1. Функције $\tilde{h}_c(x_1, \ldots, x_c)$ су симетричне по својим аргументима;
2. $\tilde{h}_c(X_1, \ldots, X_c) = E(h(X_1, \ldots, X_m)|X_1, \ldots, X_c)$;
3. $\tilde{h}_c(x_1, \ldots, x_c) = \int \tilde{h}_{c+1}(x_1, \ldots, x_{c+1}) dP(x_{c+1})$;
4. $\tilde{h}_0 \equiv \theta(P), \tilde{h}_m = h$.

Дефиниција 2.0.5 *Дефинишимо, за $0 \leqslant c \leqslant m$ функције*

$$h_c(x_1, \ldots, x_c) = \sum_{k=0}^{c} (-1)^{c-k} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant c} \tilde{h}_k(x_{i_1}, \ldots, x_{i_k}).$$

Рачунањем првих пар чланова добијамо

$$h_1(x_1) = \tilde{h}_1(x_1) - \theta$$

$$h_2(x_1, x_2) = \theta - \tilde{h}_1(x_1) - \tilde{h}_1(x_2) + \tilde{h}_2(x_1, x_2)$$

$$h_3(x_1, x_2, x_3) = -\theta + \sum_{i=1}^{3} \tilde{h}_1(x_i) + \sum_{1 \leqslant i_1 < i_2 \leqslant 3} \tilde{h}_2(x_{i_1}, x_{i_2})$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$h_m(x_1, \ldots, x_m) = (-1)^m +$$

$$+(-1)^{m-1} \sum_{i=1}^{m} \tilde{h}_1(x_i) + (-1)^{m-2} \sum_{1 \leqslant i_1 < i_2 \leqslant m} \tilde{h}_2(x_{i_1}, x_{i_2}) + \cdots + h(x_1, \ldots, x_m).$$

Приметимо да за функције h_c важе следеће особине:

1. Функције h_c су симетричне по својим аргументима;
2. Функције h_c су дегенерисане, за свако $0 \leqslant c \leqslant m$;

Резултат се добија применом особине 3. фунцкија $\tilde{h}_c$

$$E(h_1(X_1)) = 0$$

$$E(h_2(x_1, X_2)) = E(\theta - \tilde{h}_1(x_1) - \tilde{h}_1(X_2) + \tilde{h}_2(x_1, X_2)) = \theta - \tilde{h}_1(x_1) - \theta + \tilde{h}_1(x_1) = 0$$

$$\dots\dots\dots\dots\dots$$

Уопштено, израз за $E(h(x_1, \dots, x_{c-1}, X_c))$ је:

$$\sum_{k=0}^{c} (-1)^{c-k} \sum_{1 \leqslant i_1 < \dots < i_k < c, i_k \neq c} \tilde{h}_k(x_{i_1}, \dots, x_{i_k}) +$$

$$+ \sum_{k=1}^{c} (-1)^{c-k} \sum_{1 \leqslant i_1 < \dots < i_k = c} \tilde{h}_{k-1}(x_{i_1}, \dots, x_{i_{k-1}}) = 0.$$

3. $h_0 = \theta(P)$

$$h(x_1, \dots, x_m) = \sum_{c=0}^{m} \sum_{1 \leqslant i_1 < \dots < i_c \leqslant m} h_c(x_{i_1}, \dots, x_{i_c}).$$

Кренимо од леве стране једнакости

$$\sum_{c=0}^{m} \sum_{1 \leqslant i_1 < \dots < i_c \leqslant m} h_c(x_{i_1}, \dots, x_{i_c}) =$$

$$= \sum_{c=0}^{m} \sum_{1 \leqslant i_1 < \dots < i_c \leqslant m} \sum_{k=0}^{c} (-1)^{c-k} \sum_{1 \leqslant j_1 < \dots < j_k \leqslant c} \tilde{h}_k(x_{i_{j_1}}, \dots, x_{i_{j_k}})$$

$$= \sum_{k=0}^{m} \sum_{1 \leqslant j_1 < \dots < j_k \leqslant m} \tilde{h}_k(x_{j_1}, \dots, x_{j_k}) \sum_{c=k}^{m} (-1)^{c-k} \binom{m-k}{c-k}.$$

У последњој једнакости заменили смо редослед сумирања, прво смо бирали k индекса чиме смо одредили аргументе функција $\tilde{h}_k$, а затим смо бирали преосталих $c-k$ индекса, то чинимо на $\binom{m-k}{c-k}$ начина и сви они дају једнак сабирак у суми а то је $\tilde{h}_k(x_{i_{j_1}}, \dots, x_{i_{j_k}})$. Како је

$$\sum_{c=k}^{m} (-1)^{c-k} \binom{m-k}{c-k} = \sum_{c=0}^{m-k} (-1)^{c} \binom{m-k}{c} = \begin{cases} 1, & k = m \\ (1 + (-1))^{m-k}, & \text{иначе} \end{cases}.$$

У суми остаје само један сабирак

$$\sum_{c=0}^{m} \sum_{1\leqslant i_1<\cdots<i_c\leqslant m} h_c(x_{i_1},\ldots,x_{i_c}) =$$

$$= \sum_{1\leqslant j_1,\ldots,j_m\leqslant m} \tilde{h}_m(x_{j_1},\ldots,x_{j_m}) = \tilde{h}_m(x_1,\ldots,x_m) = h(x_1,\ldots,x_m),$$

чиме смо добили тражено. □

Улога функција h_c и $\tilde{h}_c$ је техничка, оне нам представљају алат за налажење израза за дисперзију U-статистика у погодном облику. Следећа теорема нас доводи корак ближе ка томе.

Теорема 2.5 *О декомпозицији U-статистика Нека је h симетрично језгро реда m и непристрасна оцена параметра θ. Тада за претходно дефинисане фунцкије h_c и свако $n \geqslant m$ важи*

$$U_n(h) = \sum_{c=0}^{m} \binom{m}{c} U_n(h_c).$$

Доказ

$$U_n(h) = \binom{n}{m}^{-1} \sum_{1\leqslant i_1<\cdots<i_m\leqslant n} \sum_{c=0}^{m} \sum_{1\leqslant k_1<\cdots<k_c\leqslant m} h_c(X_{i_{k_1}},\ldots,X_{i_{k_c}}) =$$

$$= \binom{n}{m}^{-1} \sum_{c=0}^{m} \sum_{1\leqslant i_1<\cdots<i_m\leqslant n} \sum_{1\leqslant k_1<\ldots,k_c\leqslant m} h_c(X_{i_{k_1},\ldots,X_{i_{k_c}}}).$$

Желимо да заменимо ред другој и трећој суми. Ако су фиксирани $(i_{k_1},\ldots,i_{k_c})$ и хоћемо да изаберемо преосталих $m-c$ индекса, то чинимо на $\binom{n-c}{m-c}$ начина и сви они дају један сабирак у суми а то је $h_c(x_{i_{k_1}},\ldots,x_{i_{k_c}})$. Малом изменом назива индекса ради једноставнијег записа добијамо

$$U_n(h) = \binom{n}{m}^{-1} \sum_{c=0}^{m} \binom{n-c}{m-c} \sum_{1\leqslant i_1<\cdots<i_c\leqslant n} h_c(X_{i_1},\ldots,X_{i_c}),$$

а пошто важи једнакост $\binom{n}{m}^{-1}\binom{n-c}{m-c}\binom{n}{c} = \binom{m}{c}$ имамо и

$$U_n(h) = \sum_{c=0}^{m} \binom{m}{c} U_n(h_c).$$

чиме је доказ завршен.

Приметимо да смо помоћу ове теореме U-статистику изразили као линеарну комбинацију U статистика са дегенерисаним језгром.

Сада ћемо декомпозицију написати у још једном облику, који ће бити кориснији при доказу неких неједнакости. Како је

$$U_n(h_c) = \binom{n}{c}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_c \leqslant n} h_c(X_{i_1}, \ldots, X_{i_c}).$$

Када то уврстимо у првобитну презентацију $U_n(h_c)$ добијамо

$$U_n - \theta = \sum_{c=1}^{m} \binom{m}{c} \binom{n}{c}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_c \leqslant n} h_c(X_{i_1}, \ldots, X_{i_c}).$$

Уведимо сада још једну нову ознаку

$$S_{nc} = \sum_{1 \leqslant i_1 < \cdots < i_c \leqslant n} h_c(X_{i_1}, \ldots, X_{i_c}),$$

која ће нам такође касније бити од користи. Следећа последица још једном повезује U-статистике са мартингалима.

Последица 2.0.1 *Нека је h интеграбилно, симетрично језгро реда m за параметар θ. Тада је, за свако фиксирано $1 \leqslant c \leqslant m$ низ $\{\binom{n}{m} U_n(h_c), n \geqslant c\}$ мартингал у односу на филтрацију $\{\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}, n \geqslant c\}$.*

Доказ

Јасно је да су испуњени услови а) и б) из дефиниције мартингала. Потребно је још доказати да важи в).

$$E\left(\binom{n}{c} U_n(h_c) | \mathcal{F}_k\right) = E\left(\sum_{1 \leqslant i_1 < \cdots < i_c \leqslant n} h_c(X_{i_1}, \ldots, X_{i_c}) | \mathcal{F}_k\right) =$$

$$= \sum_{1 \leqslant i_1 < \cdots < i_c \leqslant n} E(h_c(X_{i_1}, \ldots, X_{i_c}) | \mathcal{F}_k).$$

Приметимо да уколико се неки од индекса $i_1, \ldots, i_c$ не налази у скупу $\{1, \ldots, k\}$ онда је одговарајуће математичо очекивање из суме једнако 0. Претпоставимо да је то индекс i_1. Заиста применом телескопског својства и особине дегенерисаности функција h_c добијамо

$$E(h_c(X_{i_1}, \ldots, X_{i_c}) | \mathcal{F}_k) = E(E(h_c(X_{i_1}, \ldots, X_{i_c}) | X_1, \ldots, X_k, X_{i_2}, \ldots, X_{i_c}) | \mathcal{F}_k) =$$

$$= E(E(h_c(X_{i_1}, \ldots, X_{i_c})|X_{i_2}, \ldots, X_{i_c})|\mathcal{F}_k) = 0.$$

Дакле, у нашој суми ће опстати они сабирци за чије индексе важи
$1 \leqslant i_1 < \cdots < i_c \leqslant k$.

$$E\left(\binom{n}{c}U_n(h_c)|\mathcal{F}_k\right) = \sum_{1\leqslant i_1,\ldots,i_c\leqslant k} E(h_c(X_{i_1}, \ldots, X_{i_c})|\mathcal{F}_k) = \binom{k}{c}U_k(h_c),$$

□

Циљ овог поглавља је израчунавање дисперзије U-статистика. Како ће нам у томе користити лема о декомпозицији уведимо ознаке дисперзија помоћних функција.

Дефиниција 2.0.6 *Дефинишимо за $0 \leqslant c \leqslant m$ функције*

$$\zeta_c = D(\tilde{h}_c(X_1, \ldots, X_c))$$
$$\delta_c = D(h_c) = E({h_c}^2),$$

уз претпоставку да су ове дисперзије коначне.

Лема 2.0.2 *Веза између ζ_c и δ_c дата је изразима*

$$\delta_c = \sum_{i=0}^{c-1} \binom{c}{i}(-1)^i \zeta_{c-i},$$

и

$$\zeta_c = \sum_{i=0}^{c-1} \binom{c}{i}\delta_{c-i}.$$

Доказ

Из особина функција h_c видимо да је $\delta_0 = \zeta_0 = 0$. За $c \geqslant 1$ користимо већ поменуту једнакост $\sum_{k=0}^{c}(-1)^k\binom{c}{k} = 0$.

$$\delta_c = E{h_c}^2 = E\left(\sum_{k=0}^{c}(-1)^{c-k}\sum_{1\leqslant i_1<\cdots<i_k\leqslant c}(\tilde{h}_k(X_{i_1}, \ldots, X_{i_k}) - \theta)\right)^2.$$

Да не бисмо компликовали запис, приметимо да ћемо имати суму по скуповима K и H, облика $\{i_1, i_2, \ldots, i_{k_1}\}$ за $1 \leqslant i_1 < \cdots < i_{k_1} \leqslant c$, односно $\{j_1, j_2, \ldots, j_{k_2}\}$ за $1 \leqslant j_1 < \cdots < j_{k_2} \leqslant c$, где је $1 \leqslant k_1, k_2 \leqslant c$. Суму можемо записати једноставније као

$$\delta_c = \sum_{K,H\subset\{1,\dots,c\}} (-1)^{2c-|K|-|K|} E((\tilde{h}_{k_1}(X_k)-\theta)(\tilde{h}_{k_2}(X_H)-\theta)).$$

Дакле, у суми су неодређени и елеметни скупова K и H и њихов број. Приметимо да, ако су скупови K и H дисјунктни, због независности елемената у узорку имамо производ очекивања. Због једнаке расподељености елемената у узорку, вредност $E(\tilde{h}_k(X_K)-\theta)^2$ зависи само од броја елемената скупа K, не и од њихових тачних вредности. Претпоставимо да скупови K и H имају тачно l заједничких елемената. Очекивање из суме тада можемо записати као

$$E[(\tilde{h}_{k_1}(X_K)-\theta)(\tilde{h}_{k_2}(X_H)-\theta)] = E(E[(\tilde{h}_{k_1}(X_K)-\theta)(\tilde{h}_{k_2}(X_H)-\theta)|X_1,\dots,X_l]).$$

Приметимо да смо означили заједничке елементе са $X_1,\dots,X_l$, због једнакорасподељености то је могуће, а остали аргументи функција $\tilde{h}_{k_1}$ и $\tilde{h}_{k_2}$, редом $X_{l+1},\dots,X_{k_1}$ и $X'_{l+1},\dots,X'_{k_2}$ су независне величине. На основу особина условног математичког очекивања

$$\begin{aligned}
&E[(\tilde{h}_{k_1}(X_K)-\theta)(\tilde{h}_{k_2}(X_H)-\theta)|X_1,\dots,X_l] = \\
&= E[E((h(X)-\theta)(h(X)-\theta)|X_1,\dots,X_{k_1},X'_{l+1},\dots,X'_{k_2})|X_1,\dots,X_l] \\
&= E[E((h(X)-\theta)(h(X)-\theta)|X_1,\dots,X_l)|X_1,\dots,X_{k_1},X'_{l+1},\dots,X'_{k_2}] \\
&= E((\tilde{h}_l(X_1,\dots,X_l)-\theta)^2|X_1,\dots,X_{k_1},X'_{l+1},\dots,X'_{k_2}) \\
&= E(\tilde{h}_l(X_1,\dots,X_l)-\theta)^2.
\end{aligned}$$

Одавде се почетна сума своди на:

$$\begin{aligned}
\delta_c &= \sum_{K,H\subset\{1,\dots,c\}} (-1)^{|K|+|H|} E[(\tilde{h}_l(X_{K\cap H})-\theta)^2] \\
&= \sum_{i=1}^{c} D(\tilde{h}_l)\binom{c}{l}\sum_{i=0}^{c-l}\sum_{j=0}^{c-l-i}\binom{c-l}{i}\binom{c-l-i}{j}(-1)^{2l+i+j} \\
&= \sum_{i=1}^{c} D(\tilde{h}_l)\binom{c}{l}\sum_{i=0}^{c-l}\binom{c-l}{l}(-1)^i\sum_{j=0}^{c-l-i}\binom{c-l-i}{j}(-1)^j \\
&= \sum_{i=1}^{c} D(\tilde{h}_l)\binom{c}{l}(-1)^{c-l} = \sum_{i=0}^{c-1}\binom{c}{i}(-1)^i\zeta_{c-i}.
\end{aligned}$$

Други део теореме ћемо радити индукцијом по c. За $c = 1$, по дефиницији функција важи $h_1 = \tilde{h}_1 - \theta$, па је $\delta_1 = \zeta_1 \sum_{i=0}^{0} \binom{1}{i} \delta_{1-i}$. Претпоставимо да тврђење важи за свако $c = 1, \ldots, r-1$ и испитујемо да ли важи и за $c = r$. На основу претходно доказаног и применом индуктивне хипотезе, добијамо

$$\delta_r = \zeta_r + \sum_{i=1}^{r-1} \binom{r}{i} (-1)^i \zeta_{r-i} = \zeta_r + \sum_{i=1}^{r-1} \binom{r}{i} (-1)^i \sum_{j=0}^{r-i-1} \binom{r-i}{j} \delta_{r-i-j}$$

$$= \zeta_r + \sum_{i=1}^{r-1} \binom{r}{l} \delta_{r-l} \sum_{i=1}^{l} \binom{l}{i} (-1)^i = \zeta_r - \sum_{i=1}^{r-1} \binom{r}{l} \delta_{r-l}.$$

□

Последица 2.0.2 *Нека је h симетрично језгро реда m, ако је $\zeta_m < +\infty$, тада важи*

$$0 \leqslant dD\tilde{h}_c \leqslant cD\tilde{h}_d, 1 \leqslant c \leqslant d \leqslant m.$$

Доказ

$$cD\tilde{h}_d - dD\tilde{h}_c = c \sum_{i=0}^{d-1} \binom{d}{i} \delta_{d-i} - d \sum_{j=0}^{c-1} \binom{c}{j} \delta_{c-j}.$$

Груписаћемо чланове уз $\delta_0, \ldots, \delta_d$.

$$cD\tilde{h}_d - dD\tilde{h}_c = \sum_{i=1}^{c} [c \binom{d}{i} - d \binom{d}{i}] \delta_i + \sum_{i=c+1}^{d} c \binom{d}{i} \delta_i \geqslant 0.$$

Последња неједнакост се једноставно добија из чињенице да је

$$\frac{c\binom{d}{i}}{d\binom{c}{i}} > 1.$$

На основу овога имамо да важи и $0 = \zeta_0 \leqslant \cdots \leqslant \zeta_m = D(h(X_1, \ldots, X_m)) < +\infty$ □

Теорема 2.6 *Нека је h симетрично језгро реда m и нека важи*

$$\int \cdots \int h(x_1, \ldots, x_m)^2 dP(x_1) \ldots dP(x_m) < +\infty.$$

Тада је дисперзија одговарајуће U-статистике

$$DU_n(h) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c} \binom{n-m}{m-c} \delta_c.$$

Доказ

Приметимо да из датог услова теореме, на основу претходне леме, следеи да су сви $\zeta_c < +\infty, 1 \leqslant c \leqslant m$.

$$DU_n(h) = E(U_n(h) - \theta)^2 = E\left(\sum_{c=0}^{m} \binom{m}{c} U_n(h_c)\right)^2.$$

У овом математичком очекивању јављају се производи U-статистика које имају дегенерисана језгра различитих редова. Важи да, ако имамо две U-статистике различитих редова m и m', онда за n довољно велико, $n > \max(m, m')$, важи

$$E(U_n(h)U_n(h')) =$$

$$E\left(\binom{n}{m}^{-1}\binom{n}{m'}^{-1} \sum_{1\leqslant i_1<\cdots<i_m\leqslant n}\ \sum_{1\leqslant j_1<\cdots<j_{m'}\leqslant n} h(X_{i_1}, \ldots, X_{i_m})h'(X_{j_1}, \ldots, X_{j_{m'}})\right)$$

$$= \binom{n}{m}^{-1}\binom{n}{m'}^{-1} \sum_{1\leqslant i_1<\cdots<i_m\leqslant n}\ \sum_{1\leqslant j_1<\cdots<j_{m'}\leqslant n} E(h(X_{i_1}, \ldots, X_{i_m})h'(X_{j_1}, \ldots, X_{j_{m'}})).$$

Како су језгра различитих редова, то значи да постоји бар једна случајна промењлива која се јавља у тачно једном језгру. Нека је то, на пример i_m. Када интегралимо по тој промењљивој, добијамо:

$$\int h(X_{i_1}, \ldots, X_{i_m})h'(X_{j_1}, \ldots, X_{j_{m'}})dP(x_{i_m})$$

$$= h'(X_{j_1}, \ldots, X_{j_{m'}}) \int h(X_{i_1}, \ldots, X_{i_m})dP(x_{i_m}) = 0.$$

Последња једнакост важи јер је h дегенерисана.

Вратимо се сада на наш проблем.

$$DU_n(h) = \sum_{c=0}^{m} \binom{m}{c}^2 E(U_n(h_c))^2$$

$$= \sum_{c=0}^{m} \binom{m}{c}^2 E\left(\binom{n}{c}^{-1} \sum_{1\leqslant i_1<\cdots<i_c\leqslant n} h_c(X_{i_1}, \ldots, X_{i_c})\right)^2$$

$$= \sum_{c=0}^{m} \binom{m}{c}^2 \binom{n}{c}^{-2} E\left(\sum_{1\leqslant i_1<\cdots<i_c\leqslant n} h_c(X_{i_1}, \ldots, X_{i_c})\right)^2$$

$$= \sum_{c=0}^{m} \binom{m}{c}^2 \binom{n}{c}^{-2} \sum_{1\leqslant i_1<\cdots<i_c\leqslant n} \sum_{1\leqslant j_1<,\ldots,j_c\leqslant n} E(h_c(X_{i_1},\ldots,X_{i_c})h_c(X_{i_1},\ldots,X_{i_c})).$$

Уколико постоји бар један индекс, такав да се одговарајућа случајна величина налази као аргумент само у једној функцији h_c, онда је према резултату претходне леме очекивање производа тих фунцкија једнако 0.

$$\begin{aligned} DU_n(h) &= \sum_{c=0}^{m} \binom{m}{c}^2 \binom{n}{c}^{-2} \sum_{1\leqslant i_1,\ldots,i_c\leqslant n} E(h_c(X_{i_1},\ldots,X_{i_c}))^2 \\ &= \sum_{c=0}^{m} \binom{m}{c}^2 \binom{n}{c}^{-1} \delta_c = \binom{n}{m}^{-1} \sum_{c=0}^{m} \binom{m}{c}\binom{n-m}{m-c}\delta_c. \end{aligned}$$

□

Теорема 2.7 *Нека је h симетрично језгро реда m. Ако је $h_c \equiv 0$ за све $1 \leqslant c \leqslant d-1$ где је $1 \leqslant d \leqslant m$ и $\zeta_m < +\infty$, тада $n^{\frac{d}{2}}\binom{m}{d}U_n(h)$ конвергира у расподели ка некој расподели F ако и само ако $n^{\frac{d}{2}}(U_n(h)-\theta)$ конвергира ка тој истој расподели.*

Доказ теореме се може наћи у [1]

Обележимо са $Y_n = n^{\frac{d}{2}}\binom{m}{d}U_n(h_d)$ и $Z_n = n^{\frac{d}{2}}(U_n(h)-\theta)$. На основу претходне теореме имамо да $(Y_n - Z_n) \to 0$, кад $n \to +\infty$, у L_2 смислу. Тада, ако један од ова два низа конвергира у расподели, мора и други и то ка истој расподели. Као специјалан случај издвајамо $d = 1$. Тада је $\sqrt{n}mU_n(h_1)$ представљеносумом независних једнакорасподељених случајних величина $\tilde{h}_1(X_i)$, транслираних и скалираних.

$$\sqrt{n}mU_n(h_1) = \frac{m}{\sqrt{n}}\sum_{i=1}^{n} h_1(X_i) = \sqrt{n}mU_n(h_1) = \frac{m}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{h}_1(X_i)-\theta)$$

Да бисмо могли да применимо централну граничну теорему, потребно је да математичко очекивање и дисперзија буду коначни, што је испуњено, јер је

$$D(\tilde{h}_1(X_1)) = \zeta_1 - \theta^2.$$

Добијени резултат издвојићемо у следећој теореми:

Теорема 2.8 *Нека је h произвољно језгро реда m, и важи $\zeta_m < +\infty$. Тада $\sqrt{n}(U_n(h)-\theta)$ конвергира у расподели ка $\mathcal{N}(0, m^2\zeta_1)$.*

Пример 6 *Израчунајмо асимптотску дисперзију за U-статистике из примера 1 за k-ти моменат. Асимптотска нормалност важи ако је $E|X_1|^{2k} < +\infty$, а асимптотска дисперзија је одређена са DX_1^k. Слично, узорачка дисперзија као U-статистика има асимптотски нормалну расподелу $\mathcal{N}(0, 4\zeta_1)$ ако важи $E{X_1}^4 < +\infty$. У овом случају имамо следећу ситуацију:*

$$h_1(x_1) = \tilde{h}_1(x_1) - DX_1 = E(h(x_1, X_2)) - DX_1 = E\left(\frac{(x_1 - X_2)^2}{2}\right) - DX_1$$

$$= \frac{1}{2}E({x_1}^2 - 2x_1X_2 + {X_2}^2) - DX_1$$

$$= \frac{1}{2}{x_1}^2 - x_1EX_2 + \frac{1}{2}E{X_2}^2 + \frac{1}{2}(EX_2)^2 - \frac{1}{2}(EX_2)^2 - DX_1$$

$$= \frac{1}{2}(x_1 - EX_2)^2 + \frac{1}{2}(E{X_2}^2 + (EX_2)^2) - DX_1 = \frac{1}{2}(x_1 - EX_1)^2 - \frac{1}{2}DX_1.$$

Одавде следи да је

$$4\zeta_1 = Dh_1(X_1) = E(X_1 - EX_1)^4 - (DX_1)^2.$$

Ако на пример претпоставимо $X_1 \in \mathcal{N}(\mu, \sigma^2)$ тада је $4\zeta_1 = 2\sigma^4$

Пример 7 *Применићемо граничну теорему на пример 3 .*

$$W^+ = nU_n(h_1) + \binom{n}{2}U_n(h_2).$$

Постматрамо граничну вредност у расподели израза

$$\sqrt{n}\binom{n}{2}^{-1}(W^+ - EW^+).$$

Математичко очекивање овог израза је очигледно једнако нули:

$$\sqrt{n}\binom{n}{2}^{-1}\left(nU_n(h_1) + \binom{n}{2}U_n(h_2) - \frac{n(n+1)}{4}\right).$$

Како је

$$D(\sqrt{n}\binom{n}{2}^{-1}n(U_n(h_1) - E(U_n(h_2)))) = \binom{n}{2}^{-2}n^2Dh_1 \to 0, n \to 0,$$

значи да та случајна променљива тежи дегенерисаној. На други део израза применимо теорему 2.8

$$\sqrt{n}(U_n(h_2) - E(U_n(h_2))) \to \mathcal{N}(0, 4\sigma^2),$$

где је $\sigma^2 = D(\tilde{h}_2)_1$. *На основу овог, цео израз тежи у расподели ка* $\mathcal{N}(0, 4\sigma^2)$ *расподели. Како је*

$$(\tilde{h}_2)_1(z_1) = E(I(z_1 + Z_1 > 0)) = P(Z_2 > -z_1) = 1 - F(-z_1),$$

и, при H_0 *случајна величина* Z_1 *расподељена симетрично око нуле, следи да* $(\tilde{h}_2)_1(z_1)$ *има* $U[0,1]$ *расподелу. Одавде имамо да је* $\sigma^2 = \frac{1}{2}$. *На основу овог видимо да гранична расподела не зависи од непознатог параметра* θ, *и да је дисперзија Вилкоксонове статистике* $\frac{n(n-1)^2}{12}$.

Поглавље 3

Неједнакости за U-статистике

Увели смо основне појмове и показали везу између мартингала и U-статистика. У трећој глави ћемо их искористити за доказивање неједнакости са моментима и максималне неједнакости, које уједно чине и најважније неједнакости за U-статистике.

3.1 Неједнакости са моментима

Нека је $(S_n, \mathcal{F}_n), n = 1, 2, ...,$ случајан низ.

Теорема 3.1 *Нека су ξ_i случајне величине које задовољавају услове*

$$E(\xi_i | S_{i-1}) = 0, 2 \leqslant i \leqslant n,$$

где је $\xi_1 = S_1, \xi_i = S_i - S_{i-1}, i = 2, \ldots, n$. Тада важи

$$E|S_n|^p \leqslant \alpha_p \sum_{j=1}^{n} E|\xi_i|^p,$$

за $1 \leqslant p \leqslant 2$, где је

$$\alpha_p = \sup_x (|x|^{-p}(|1+x|^p - 1 - px)) \leqslant 2^{2-p}.$$

Доказ

Ако је $(S_i, \mathcal{F}_i)$ мартингал, тада низ $(\xi_i, \mathcal{F}_i)$ представља мартингалну разлику. Важи $S_n = \sum_{i=1}^{n} \xi_i$. За $p = 1$ добијамо $\alpha_p = 2$. Дакле тражена неједнакост постаје

$$E|\sum \xi_i| \leqslant \sum E|\xi_i|,$$

што је задовољено. За $p = 2$ имамо $ES_n{}^2 = E\xi_1{}^2 + \cdots + E\xi_n{}^2$ због услова које испуњавају величине ξ_i. Размотримо сада случај $1 < p < 2$. Користићемо елементарну неједнакост

$$|a+b|^p \leqslant |a|^p + p|a|^{p-1}\mathrm{sgn}(a)b + \alpha_p|b|^p,$$

која важи за све реалне a и b. Приметимо да $\alpha_p > 1$ за $1 < p < 2$. Ако у поменутој неједнакости уврстимо $a = \xi_1$, $b = \xi_2$ и затим интегралимо обе стране неједнакости добијамо

$$E|\xi_1 + \xi_2|^p \leqslant E|\xi_1|^p + \alpha_p E|\xi_2|^p.$$

Индукцијом добијамо тражену неједнакост, ако за индуктивну хипотезу уврстимо $a = S_{n-1}, b = \xi_n$. □

Следећа теорема такође даје горњу границу p-тог момента парцијалних мартингалних сума, а доказ се може наћи у [5].

Теорема 3.2 *Нека је $(S_n, \mathcal{F}_n)$ мартингал. Дефинишемо*

$$\gamma_{pn} = E|\xi_n|^p, \Gamma_{pn} = \frac{1}{n}\sum_{j=1}^{n}\gamma_{pj}.$$

Тада неједнакост

$$E|S_n|^p \leqslant \gamma_p n^{p/2}\Gamma_{pn}$$

важи за свако $p \geqslant 2$ и $n = 1, 2, \ldots$ где је

$$\gamma_p = (8(p-1)max(1, 2^{p-3}))^p.$$

Пример 8 *Нека је Y_n случајна величина, мерљива у односу на σ-алгебру $\mathcal{F}_n, n = 1, 2, \ldots$ таква да важи $E|Y_n| < +\infty$. Уводимо ознаку*

$$\xi_{nk} = E(Y_n|\mathcal{F}_n) - E(Y_n|\mathcal{F}_{n-1}), k = 1, \ldots, n,$$

из које добијамо

$$Y_n - EY_n = \xi_{n1} + \cdots + \xi_{nn},$$

где низ $(\xi_{nk}, \mathcal{F}_n), k \geqslant 1$ чини мартингалску разлику за свако n.

Нека је $p \geqslant 1$ фиксиран реалан број. Претпоставимо да $E|h|^p < +\infty$. Под овим условом, можемо оценити математичко очекивање $E|U_n - \theta|^p$, користећи мартингалска својства U_n. Запишимо U_n као у примеру 8. Имамо

$$U_n - \theta = \xi_{n1} + \cdots + \xi_{nn},$$

где су
$$\xi_{nk} = E(U_n|X_1, \ldots, X_k) - E(U_n|X_1, \ldots, X_{k-1}).$$
Како је низ S_{nk} мартингал (из последице 2.0.1), искористићемо алтернативни облик декомпозиције функција изведен у 3. глави. Добијамо
$$\xi_{nk} = \sum_{c=1}^{m} \binom{m}{c} \binom{n}{c}^{-1} (S_{kc} - S_{k-1,c}).$$
За $k \geqslant c, k = 1, \ldots, n, c = 1, \ldots, m$ уводимо ознаку
$$\eta_{kc} = \sum_{1 \leqslant i_1 < \cdots < i_{c-1} \leqslant k-1} h_c(X_{i_1}, \ldots, X_{i_{c-1}}, X_k).$$
По дефиницији имамо да је
$$\eta_{k1} = h_1(X_k), k = 1, 2, \ldots, n.$$
Како је
$$S_{kc} - S_{k-1,c} = \eta_{kc},$$
имамо
$$\xi_{nk} = \sum_{c=1}^{m} \binom{m}{c} \binom{n}{c}^{-1} \eta_{kc},$$
односно $(\xi_{nk}, \mathcal{F}_k), k \geqslant 1$ формирају мартингалске разлике, одакле добијамо
$$E(\xi_{n,k+1}|\xi_{n1} + \cdots + \xi_{nk}) = 0, k = 1, 2, \ldots$$
Случајне величине $\xi_{ni}, i = 1, \ldots, k+1$ задовољавају услове теореме 3.1, па користећи поменуту теорему добијамо
$$E|U_n - \theta|^p \leqslant \alpha_p \sum_{k=1}^{n} E|\xi_{nk}|^p,$$
за $1 \leqslant p \leqslant 2$. Прво ћемо оценити математичко оцекивање $E|\xi_{nk}|^p$. Имамо
$$E|\xi_{nk}|^p \leqslant (m - r + 1)^{p-1} \sum_{c=r}^{m} \binom{m}{c}^p \binom{n}{c}^{-p} E|\eta_{kc}|^p,$$
где $r \geqslant 1$ природан број представља ред језгра h. С обзиром да функције h_c имају својство дегенерисаности, низ $\{\eta_{kc}\}, k = 1, 2, \ldots$ задовољава услове теореме 3.1, па поново користећи ту теорему добијамо
$$E|\eta_{kc}|^p \leqslant {\alpha_p}^c \binom{k-1}{c-1} E|h_c|^p.$$

Напоменимо да је

$$\sum_{k=1}^{n} \binom{k-1}{c-1} = \binom{n}{c}.$$

Користећи претходне две неједнакости добијамо

$$\sum_{k=1}^{n} E|\xi_{nk}|^p \leqslant (m-r+1)^{p-1} \sum_{c=r}^{m} \binom{m}{c}^p \binom{n}{c}^{-p+1} {\alpha_p}^c E|h_c|^p,$$

за $1 \leqslant p \leqslant 2$.

Теорема 3.3 *Претпоставимо да је $Eh = 0$ и $E|h|^p < +\infty$ за свако $1 \leqslant p \leqslant 2$. Тада*

$$E|U_n - \theta|^p \leqslant (m-r+1)^{p-1} \sum_{c=r}^{m} \binom{m}{c}^p \binom{n}{c}^{-p+1} {\alpha_p}^{c+1} E|h_c|^p.$$

Размотримо сада случај $p \geqslant 2$. Према теореми 3.2 имамо

$$E|U_n - \theta|^p \leqslant {\gamma_p}^{(p-2)/2} \sum_{k=1}^{n} E|\xi_{nk}|^p.$$

Очигледно је да

$$E|\xi_{nk}|^p \leqslant (m-r+1)^{p-1} \sum_{c=r}^{m} \binom{m}{c}^p \binom{n}{c}^{-p} E|\eta_{kc}|^p,$$

такође важи и за $p \geqslant 2$. Стога, требало би оценити математичко очекивање $E|\eta_{kc}|^p$. Да бисмо то постигли, применићемо претходну теорему $c-1$ пута. Добијамо оцену

$$E|\eta_{kc}|^p \leqslant {\gamma_p}^c n^{(p-2)(c-1)/2} \binom{k-1}{c-1} E|g_c|^p.$$

Сада можемо формулисати још једну теорему чији се доказ налази у [5].

Теорема 3.4 *Претпоставимо да је $Eh = 0$ и $E|h|^p < +\infty$ за свако $p \geqslant 2$. Тада важи следећа неједнакост*

$$E|Un - \theta|^p \leqslant (m-r+1)^{p-1} \sum_{c=r}^{m} \binom{m}{c}^p \binom{n}{c}^{-p+1} n^{p(p-2)c/2} {\gamma_p}^{c+1} E|h_c|^p.$$

3.2 Максималне неједнакости

Теорема 3.5 *Претпоставимо да је $(X_n, \mathcal{F}_n)$ субмартингал, $\{a_n\}, a_n > 0$ је растући низ $\mathcal{F}_{n-1}$-мерљивих случајних величина, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ и $t > 0$. Тада*

$$tP\left(\max_{1\leqslant i\leqslant n}(X_i/a_i) \geqslant t\right) \leqslant E(X_1^+/a_1) + \sum_{i=2}^{n} E((X_i^+ - X_{i-1}^+/a_i),$$

за произвољно $n \geqslant 1$, где је $X_i^+ = \max(X_i, 0)$.

Доказ Приметимо да важи

$$P(\max_{1\leqslant i\leqslant n}(X_i/a_i) \geqslant t) = P(\max_{1\leqslant i\leqslant n}(X_i^+/a_i) \geqslant t),$$

где је $(X_i^+, \mathcal{F}_i)$ субмартингал. Без умањења општости, можемо претпоставити да је $X_i \geqslant 0$ (скоро сигурно) за свако $i \geqslant 1$. Нека је

$$Y_k = X_1/a_1 + \sum_{i=2}^{k}(X_i - X_{i-1})/a_i.$$

Низ $(Y_k, \mathcal{F}_k)$ је позитивни субмартингал. Означимо са m најмањи међу бројевима $(k = 1, 2, \ldots, n)$ који задовољава $X_k/a_k \geqslant t$, ако такав број постоји. У супротном претпоставимо да је $m = n + 1$. Можемо записати Y_k у следећем облику

$$Y_k = X_k/a_k + \sum_{i=1}^{k-1}(a_i^{-1} - a_{i+1}^{-1})/X_i.$$

На скупу $\{\omega : m = k\}$ важи

$$t \leqslant X_k/a_k \leqslant Y_k.$$

Стога, увек имамо Y_m на скупу $\{\omega : m \leqslant n\}$. Односно

$$tP\left(\max_{1\leqslant i\leqslant n}(X_i/a_i) \geqslant t\right) \leqslant tP(m \leqslant n) \leqslant E\left(Y_m I(m \leqslant n)\right) =$$

$$= \sum_{i=1}^{n}\int_{\{\omega:m=i\}} Y_i dP \leqslant \sum_{i=1}^{n}\int_{\{\omega:m=i\}} Y_n dP = E(Y_m I(m \leqslant n)) \leqslant EY_n.$$

Што је и требало доказати.

Наредне две теореме представљају последице теореме 3.5 и доказ се може наћи у [5].

Теорема 3.6 *Нека је $(X_n, \mathcal{F}_n)$ субмартингал и нека је $\{c_n\}, n \geqslant 1$ нерастући низ ненегативних бројева. Тада важи*

$$P\left(\max_{1\leqslant i\leqslant n} c_k X_k \geqslant t\right) \leqslant t^{-1}\left(c_n E X_n^+ + \sum_{i=1}^{n-1}(c_i - c_{i+1}) E X_i^+\right),$$

за $t > 0$. Посебно,

$$P\left(\max_{1\leqslant i\leqslant n} X_k \geqslant t\right) \leqslant t^{-1} E X_n^+.$$

Теорема 3.7 *Нека је $(X_n, \mathcal{F}_n)$ обрнути субмартингал, а $\{c_n\}, n \geqslant 1$, неопадајући низ позитивних бројева. Тада неједнакост*

$$P\left(\max_{n\leqslant k\leqslant N} c_k X_k \geqslant t\right) \leqslant t^{-1}\left(c_n E X_n^+ + \sum_{i=n+1}^{N}(c_i - c_{i-1}) E X_i^+)\right),$$

важи за свако $t > 0$ и $N > n$. Посебно,

$$P\left(\sup_{k\geqslant n} X_k \geqslant t\right) \leqslant t^{-1} E X_n^+.$$

Теорема 3.8 *Претпоставимо да важи $Eh^2 < +\infty$ и нека је ранг језгра h једнак $r, 1 \leqslant r \leqslant m$. Тада неједнакост*

$$P\left(\max_{m\leqslant k\leqslant n} \binom{k}{m} a_k |U_k - \theta| \geqslant t\right) \leqslant 2^{2m} t^{-2} \sum_{c=r}^{m} \delta_c c \sum_{k=c}^{n} \binom{k}{c}^{-1} k a_k^2,$$

важи за произвољно $t > 0$ и нерастући низ ненегативних бројева a_n.

Доказ Према последици 2.0.1 низови $\{S_{nc}, n = 1, 2, \dots\}, c = r, ..., m$, су мартингали. Даље следи да су ${S_{nc}}^2$ субмартингали. Користећи теорему 3.6 добијамо

$$P\left(\max_{c\leqslant k\leqslant n} a_k |S_{kc}| \geqslant t\right) = P\left(\max_{c\leqslant k\leqslant n} a_k^2 S_{kc}^2 \geqslant t^2\right)$$

$$\leqslant t^{-2}\left(\sum_{k=c}^{n-1}(a_k^2 - a_{k+1}^2) E S_{kc}^2 + a_n^2 E S_{nc}^2\right)$$

$$= t^{-2}\left(\sum_{k=c}^{n-1}(a_k^2 - a_{k+1}^2)\binom{k}{c}\delta_c + a_n^2 \binom{n}{c}\delta_c\right)$$

$$= t^{-2}\delta_c \left(\binom{c}{c} a_c^2 + \frac{c}{c+1}\binom{c+1}{c} a_{c+1}^2 + \cdots + \binom{n}{c} a_n^2 \right).$$

Сада, за $t > 0$ и свако $c = r, \ldots, m$ имамо

$$P\left(\max_{c\leqslant k\leqslant n} a_k |S_{kc}| \geqslant t \right) \leqslant t^{-2}\delta_c c \sum_{k=c}^{n} \binom{k}{c} k^{-1} a_k^2.$$

Употребићемо теорему о декомпозицији U-статистика, али пре тога уводимо догађаје

$$A = \left\{ \omega : \max_{m\leqslant k\leqslant n} a_k \binom{k}{m} |U_k - \theta| \geqslant t \right\}$$

$$A_c = \left\{ \omega : \max_{c\leqslant k\leqslant n} a_k |S_{kc}| \geqslant t_m \right\},$$

где је $t_m = 2^{-m}t$ и $c = r, \ldots, m$. Ако се сваки од догађаја $\overline{A}_r, \ldots \overline{A}_m$ догоди, тада

$$a_k \binom{k}{m} |U_k - \theta| \leqslant a_k \binom{k}{m} \sum_{c=1}^{m} S_{kc} \binom{m}{c} / \binom{k}{m} < 2^{-m} t \sum_{c=1}^{m} \binom{m}{c} = t.$$

Због тога, $A \subset \cup_{c=r}^{m} A_c$. Па имамо

$$P(A) \leqslant \sum_{c=r}^{m} P(A_c) \leqslant \sum_{c=r}^{m} (2^{-m}t)^{-2} \delta_c c \sum_{k=c}^{n} \binom{k}{c} k^{-1} a_k^2,$$

што доказује наведену неједнакост.

Теорема 3.9 *Нека је $\{c_n\}, n \geqslant m$ неопадајући низ позитивних бројева и важи $Eh^2 < +\infty$ Тада*

$$P\left(\max_{n\leqslant k\leqslant N} c_k |U_k| > t \right) \leqslant t^{-2} \left[c_n^2 \sigma^2(U_n) + \sum_{k=n+1}^{N} (c_k^2 - c_{k-1}^2) \sigma^2(U_k) \right],$$

за свако $m \leqslant n \leqslant N$ и $t > 0$. Ако десна страна израза конвергира (за $N \to \infty$), онда у овој неједнакости, $\max\limits_{n\leqslant k\leqslant N}$ и N могу бити замењени са $\sup\limits_{k\geqslant n}$, односно ∞.

Доказ По последици 2.0.1 и Јенсеновој неједнакости, низ $(U_n^2, \mathcal{B}_n), n \geqslant m$, је ненеагативни обнути субмартингал. Стога, неједнакост следи из теореме 3.7.

Поглавље 4

Примена неједнакости на доказ централне граничне теореме

Посматрамо U-статистику

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_m} h(X_{i_1}, \ldots, X_{i_m}),$$

где су $X_1, \ldots, X_n$ независне једнакорасподељене случајне величине, а језгро је симетрично и недегенерисано.

Теорема 4.1 *Ако је а) $\eta_1 > 0$ и б) $Eh^2(X_1, \ldots, X_m) < +\infty$, тада*

$$n^{1/2}(m^2\eta_1)^{-1/2}(U_n - \theta) \xrightarrow{d} \tau,$$

за $n \to +\infty$, где τ представља стандардну нормалну расподелу $\mathcal{N}(0,1)$.

Доказ Према теореми о декомпозицији (теорема 2.5) имамо

$$U_n - \theta = mn^{-1} \sum_{j=1}^{n} g_1(X_j) + R_n,$$

где је

$$R_n = \sum_{c=2}^{m} \binom{m}{c} U_n(h_c).$$

Под условом б), имамо

$$ER_n^2 = \sum_{c=2}^{m} \binom{m}{c}^2 \binom{n}{c}^{-1} Eh_c^2.$$

Дакле,

$$n^{1/2} R_n \xrightarrow{P} 0, n \to +\infty$$

Дато тврђење следи из чињенице да се у расписаном облику U_n статистике налази сума $h_1(X_1) + \cdots + g_1(X_n)$ независних једанко расподељених случајних величина, које имају позитивну дисперзију због услова а). Тиме је доказ завршен.

Теорема 4.2 *Претпоставимо да језгро h има ранг $r = 1$ и задовољава услов в)*

$$E|h_c|^{\gamma_c} < +\infty, c = 1, ..., m,$$

где је $\gamma_c = 2c/(2c-1)$, тада поново важи

$$n^{1/2} (m^2 \eta_1)^{-1/2} (U_n - \theta) \xrightarrow{d} \tau.$$

Доказ За почетак приметимо

$$2 = \gamma_1 > \gamma_2 > \cdots > \gamma_m = 1 + \frac{1}{2m-1} > 1.$$

Такође услов в) је слабији него услов б). За $c = 1$, из услова в) следи да $Eh_1^2(X_1) < +\infty$. Сада желимо да покажемо да $n^{1/2} R_n \xrightarrow{P} 0, n \to +\infty$ важи под условом в) за $2 \leqslant c \leqslant m$. Како бисмо то урадили представићемо $U_n(h_c)$ за свако $c = 2, \ldots, m$ као суму две U-статистике

$$U_{nc} = U_n(\varphi_{1c}) + U_n(\varphi_{2c}),$$

са језгрима

$$\varphi_{kc}(x_1, \ldots, x_c) = \sum_{d=1}^{c} (-1)^{c-d} \sum_{1 \leqslant j_1 < \cdots < j_d \leqslant c} \tilde{h}_{cd}^{(k)}(x_{j_1}, \ldots, x_{j_d}), k = 1, 2,$$

где

$$\tilde{h}_{cd}^{(k)}(x_1, \ldots, x_d) = h_{cd}^{(k)}(x_1, \ldots, x_d) - Eh_{cd}^{(k)}(X_1, \ldots, X_d)$$

$$h_{cd}^{(1)}(x_1, \ldots, x_d) =$$

$$= E(h_c(X_1,\ldots,X_c)I(|h_c(X_1,\ldots,X_c)| \leqslant \sqrt{n})|X_1 = x_1,\ldots,X_d = x_d)$$

$$h_{cd}^{(2)}(x_1,\ldots,x_d) =$$
$$= E(h_c(X_1,\ldots,X_c)I(|h_c(X_1,\ldots,X_c)| > \sqrt{n})|X_1 = x_1,\ldots,X_d = x_d).$$

$I(A)$ представља индикатор скупа A. Приметимо да важи $h_c = \varphi_{1c} + \varphi_{2c}$ и функције $\varphi_{1c}, \varphi_{2c}$ су потуно дегенерисане. Узимајући ово у обзир добијамо

$$EU_n^2(\varphi_{1c}) = \binom{n}{c}^{-1} E\varphi_{1c}^2 \leqslant \binom{n}{c}^{-1} 2^{2c} E(h_c^2(X_1,\ldots,X_c)I(|h_c(X_1,\ldots,X_c)| \leqslant \sqrt{n}))$$

$$= \binom{n}{c}^{-1} 2^{2c} E(|h_c|^{\gamma_c}|h_c|^{2-\gamma_c} I(|h_c| < \sqrt{n})) \leqslant 2^{2c}\binom{n}{c}^{-1} n^{1-\gamma_c/2} E|h_c|^{\gamma_c}.$$

Одакле имамо

$$E|\sqrt{n}U_n(\varphi_{2c}))|^{\gamma_c} = O(n^{-(\gamma_c/2)^2}),$$

за $c = 2,\ldots,m$

За $U_n(\varphi_{2c})$, користећи теорему 3.3 добијамо

$$E|U_n(\varphi_2 c)|^{\gamma_c} \leqslant 2^{c+1}\binom{n}{c}^{-\gamma_c+1} E|\varphi_{2c}|^{\gamma_c}$$

$$\leqslant 2^{3c+1}\binom{n}{c}^{-\gamma_c+1} E(|h_c|^{\gamma_c} I(|h_c| > \sqrt{n})).$$

Дакле

$$E|\sqrt{n}U_n(\varphi_{2c})|^{\gamma_c} \leqslant 2^{3c+1}\left(c!(1-\frac{1}{n})^{-1}\ldots(1-\frac{c}{n})^{-1}\right)^{-\gamma_c+1} E(|h_c|^{\gamma_c} I(|h_c| > \sqrt{n})),$$

за $c = 2,\ldots,m$. Даље следи

$$E|\sqrt{n}U_n(\varphi_2 c)|^{\gamma_c} \to 0, n \to +\infty,$$

за $c = 2,\ldots,m$ ако важи услов в). У складу са дефиницијиом R_n, за $\varepsilon > 0$ имамо

$$P(|\sqrt{n}R_n| > \varepsilon) \leqslant \sum_{c=2}^{m} P\left((m-1)\binom{m}{c}|\sqrt{n}U_{nc}| > \varepsilon\right)$$

$$\leqslant \sum_{c=2}^{m} P\left(2(m-1)\binom{m}{c}|\sqrt{n}U_n(\varphi_{1c})| > \varepsilon\right)$$
$$+\sum_{c=2}^{m} P\left(2(m-1)\binom{m}{c}|\sqrt{n}U_n(\varphi_{2c})| > \varepsilon\right).$$

Напокон користећи Чебишевљеву неједнакост добијамо

$$P(|\sqrt{n}R_n| > \varepsilon) \to 0, n \to +\infty.$$

Дакле доказали смо теорему под слабијим условима.

Закључак

Фински статистичар Васили Хоефдинг (1914-1991) је био један од првих статистичара који се бавио непараметарским статистикама, а највише је допринео тој области тако што је увео појам и извео неке прве резултате везане за U-статистике. Испитивање неке уопштене класе статистика, а не појединачних статистика, погодно је због општих резултата које тиме добијамо. На примерима у првом поглављу смо показали да класа U-статистика садржи неке ве познате статистике. У другом поглављу смо приближили богату теорију мартингала и одатле извесли неке особине U-статистика као оцена параметара. Нашли смо израз за дисперзију U-статистика, који су нам биле од користи у доказу граничних теорема у трећем Поглављу. Резултате доказане трећем поглављу смо искористили за доказ Централне граничне теореме.

Литература

[1] Denker, Manfred. (1985). Asymphtotic Distibution Theory in Nonparametric Statistics. Springer Fachmedien Wiesbaden GmbH. pp 1-50

[2] Serfling,Robert. (1980). Approximation Theorems of Mathematical Statistics. New York, NY:Wiley. pp. 171-209

[3] Ferguson, Thomas.(2005). U-statistics (Notes for Statistics 200C)

[4] Mladenovic, Pavle. (2005.) Verovatnoca i statistika. Matematicki fakultet, Beograd.

[5] Koroljuk V.S, Borovskich Yu.V. (1994). -Theory of U-Statistics-Springer Netherlands (Mathematics and Its Applications 273).

Биографија

Рођен у Београду, 6. априла 1992. године. Првих шест разреда основне школе похађао у ОШ Бора Станковић на Вождовцу. Седми и осми разред, као и средњошколско образовање стекао у Математичкој гимназији у Београду. Математички факултет, уписао 2011. године и дипломирао 2015. године. Исте године уписује мастер студије на смеру Статистика, актуарска и финансијска математика.