A spektrum becslése

A spektrál-sűrűségfüggvényt becsüljük, ha létezik, illetve tiszta diszkrét spektrum esetén a frekvenciákat és az "ugrások" nagyságát. Harmonikus folyamat esetén, ha $X(t) = \sum\limits_{i=1}^{K} A_i \cos(\lambda_i t + \phi_i(\omega))$, A_i-k és λ_i-k konstansok, és a ϕ_i fázis egyenletes eloszlású valószínűségi változó $(-\pi, \pi)$-n - persze egy rögzített ω-ra $\phi_i(\omega)$ konstans[1], és így egy periodikus függvényből kell és lehet matematikailag meghatározni a paramétereket, ami nem statisztikai feladat.

Egy "realisztikusabb" modellt kapunk, ha a megfigyeléshez hozzáadódó zajt is figyelembe vesszük, azaz

$$X(t) = \left[\sum_{i=1}^{K} A_i \cdot \cos(\lambda_i t + \phi_i(\omega))\right] + \varepsilon_t.$$

Az ismeretlen paraméterek $\kappa, A_i, \lambda_i, \sigma_\varepsilon^2$ $(i = 1, \ldots, K)$. Ezt a cos-ra vonatkozó addíciós tétellel másképp is felírhatjuk:

$$X(t) = \sum_{i=1}^{K} (a_i \cdot \cos \lambda_i t + b_i \cdot \sin \lambda_i t) + \varepsilon_t,$$

ahol $a_i = A_i \cos \phi_i$, $b_i = A_i \sin \phi_i$. Innen persze $A_i = \sqrt{a_i^2 + b_i^2}$ és $\phi_i = \operatorname{arctg}\left(-\frac{b_i}{a_i}\right)$. Tételezzük fel, hogy λ_i-k **a priori ismertek** és a_i-t, b_i-t tekintsük a rendszer paramétereinek[2]. Ezek legkisebb négyzetes becslése a

$$Q = \sum_{t=1}^{N} \left\{ X_t - \sum_{i=1}^{K} a_i \cdot \cos(\lambda_i t) + b_i \cdot \sin(\lambda_i t) \right\}^2$$

kifejezés minimalizálását - tulajdonképpen egy lineáris regressziós feladat legkisebb négyzetes megoldását - jelenti. Ha deriváljuk Q-t a_i, b_i függvényében, a

$$\begin{cases} \sum\limits_{t=1}^{N} X_t \cdot \cos(\lambda_j t) = \sum\limits_{i=1}^{K} \hat{a}_i \cdot c_{i,j} + \sum\limits_{i=1}^{K} \hat{b}_i \cdot d_{i,j} \\ \sum\limits_{t=1}^{N} X_t \cdot \sin(\lambda_j t) = \sum\limits_{i=1}^{K} \hat{a}_i \cdot d_{i,j} + \sum\limits_{i=1}^{K} \hat{b}_i \cdot s_{i,j} \end{cases} \qquad j = 1, \ldots, K$$

egyenletrendszerhez jutunk, ahol

$$c_{i,j} = \sum_{t=1}^{N} \cos(\lambda_i t) \cdot \cos(\lambda_j t), \quad s_{i,j} = \sum_{t=1}^{N} \sin(\lambda_i t) \cdot \sin(\lambda_j t), \quad d_{i,j} = \sum_{t=1}^{N} \sin(\lambda_i t) \cdot \cos(\lambda_j t).$$

Ez a bonyolult egyenletrendszer jelentősen leegyszerűsödik, ha az összes λ_i frekvencia a $\frac{2\pi}{N}$ egész számú többszöröse, más szóval ha N a periódusok egész számú többszöröse. Ekkor ugyanis érvényes sin és cos függvények ortogonalitási relációja, azaz $i \neq j$ esetén $c_{i,j} = s_{i,j} = 0$, és $d_{i,j} = 0$ bármely i, j-re. Ha $p_i \neq 0$ és $p_i \neq \frac{N}{2}$, amikor N páros, akkor $c_{i,i} = s_{i,i} = \frac{N}{2}$. Ha $p_i = 0$ vagy $\frac{N}{2}$ amikor N páros, akkor $c_{i,i} = N$ és $s_{i,i} = 0$. Így tehát, ha $\lambda_i = \frac{2\pi}{N} \cdot p_i$ valamely $0 < p_i < \left[\frac{N}{2}\right]$ egészekre, akkor az egyenleteinkből

$$\hat{a}_i = \frac{2}{N} \cdot \sum_{t=1}^{N} X_t \cdot \cos(\lambda_i t), \quad \hat{b}_i = \frac{2}{N} \cdot \sum_{t=1}^{N} X_t \cdot \sin(\lambda_i t).$$

Ha például $p_i = 0$, akkor $c_{i,i} = N$ miatt $\hat{a}_i = \frac{1}{N} \cdot \sum\limits_{t=1}^{N} X_t \cdot \cos(\lambda_i t) = \overline{X}$ mintaátlagot kapjuk. A többi speciális eset hasonlóan adódik.

[1] így az életben pontosan ilyen modellek nem igazán fordulnak elő

[2] ha a_i-t és b_i-t véletlen számoknak tekintjük, akkor nyugodtan feltételezhetjük róluk, hogy normális eloszlásból származnak, hisz két normális hányadosa Cauchy, aminek az arctg -e egyenletes lesz, ami éppen ϕ_i eloszlása

Az ismeretlen frekvenciák esete

Ha a λ_i-k nem ismertek, akkor az $X(t)$-re vonatkozó "lineáris feltétel", s így a standard legkisebb négyzetes módszert sem tudjuk alkalmazni. A frekvenciák keresése az ún. periodogram függvénnyel lehetséges, melyet Schuster vezetett be először 1898-ban. A módszer heurisztikája a következő. Tegyük fel, hogy megtippeljük λ_1 értékét, legyen ez mondjuk $\hat{\lambda}_1$, majd ezt behelyettesítve az ismert λ_i-k esetén kiszámolt képletbe, elkészítjük az $\hat{a}_1, \hat{b}_1$ becsléseket. Ha jól tippeltük meg λ_1 értékét, akkor a becslések közel lesznek az igazi a_1, b_1-hez, így a becsült $\hat{A}_1^2 = \hat{a}_1^2 + \hat{b}_1^2$ amplitúdónégyzet nem lesz 0, mivel közel lesz az igazi A_i^2-hez. Ha azonban a $\hat{\lambda}_1$ "messze" van a valódi λ_1-től (és a többi frekvenciától is), akkor egy olyan tag együtthatóját becsüljük, ami nincs is benne a modellben, azaz $\hat{A}_1^2$ a $\hat{\lambda}_1$ frekvenciához tartozó amplitúdónégyzetet, azaz a 0-t becsüli, tehát a becslés 0 közeli lesz. Hogy ezt a hatást erősítsük, az amplitúdónégyzet helyett annak $\frac{N}{2}$-szeresét becsüljük, így jobban kiugranak a "csúcsok", más szóval azok a frekvenciák, amelyek valóban szerepelnek a modellben. Vezessük be tehát a következő definíciót:

5.1. Definíció. *Minden $\lambda \in (-\pi, \pi)$-re legyen az $I_N(\lambda) = (a(\lambda))^2 + (b(\lambda))^2$ függvény az ún.* ***periodogram függvény****, ahol $a(\lambda) = \sqrt{\frac{2}{N}} \sum_{t=1}^{N} X_t \cdot \cos(\lambda t)$ és $b(\lambda) = \sqrt{\frac{2}{N}} \sum_{t=1}^{N} X_t \cdot \sin(\lambda t)$.*

Másképpen írva $I_N(\lambda) = \frac{2}{N} \left| \sum_{t=1}^{N} X_t \cdot e^{-i\lambda t} \right|^2$. [3]

Ezt a függvényt persze minden λ-ra nem tudjuk kiszámolni, csak egy diszkrét frekvenciakészletre. Nevezetesen a $\lambda_p = 2\pi \cdot \frac{p}{N}$ $(p = 0, 1, \dots, \left[\frac{N}{2}\right])$ frekvenciákra[4] szokás kiszámítani. Ha ez a λ_p éppen a modellezett folyamat valamely λ_i frekvenciájával esik egybe, akkor $a(\lambda_p) = \sqrt{\frac{N}{2}} \cdot \hat{a}_i$, illetve $b(\lambda_p) = \sqrt{\frac{N}{2}} \cdot \hat{b}_i$.

Ezért ekkor $EI_N(\lambda_p) \overset{jel.}{=} E(I_p) = \frac{N}{2}\left[E\left(\hat{a}_i^2\right) + E(\hat{b}_i^2)\right]$, ahol a zárójelen belüli tagok $E\hat{a}_i^2 = (E\hat{a}_i)^2 + D^2(\hat{a}_i) = E(\hat{a}_i)^2 + D^2\left(\frac{2}{N} \sum_{t=1}^{N} X_t \cos(\lambda_i t)\right) = E(\hat{a}_i)^2 + \left(\frac{2}{N}\right)^2 \sigma_\varepsilon^2 \sum_{t=1}^{N} \cos^2(\lambda_i t) = a_i^2 + \frac{2\sigma_\varepsilon}{N}$ (b_i-re ugyanígy). Az utolsó lépésben felhasználtuk a cos-ra vonatkozó ortogonalitási relációt. Ez utóbbit behelyettesítve

$$E(I_p) = \frac{N}{2}(a_i^2 + b_i^2) + 2 \cdot \sigma_\varepsilon^2 = \frac{N}{2} \cdot A_i^2 + 2 \cdot \sigma_\varepsilon^2$$

Ha $\lambda_p \neq \lambda_i$, akkor vezessük be λ_{K+1}-ként λ_p-t, 0 amplitúdóval, azaz $A_{K+1} = 0$, és így a fenti alkalmazható, méghozzá $E(I_p) = 2 \cdot \sigma_\varepsilon^2$. Tehát, ha λ_p valamely λ_i közelében van, akkor a periodogram $O(N)$ nagyságrendű, míg ha λ_p távol van λ_i-ktől, a nagyságrend $O(1)$. Az igazi frekvenciákat tehát a csúcsok helyével határozhatjuk meg, ami persze számos további kérdést felvet. Például mi lesz akkor, ha λ_p közel van, de nem egyenlő λ_i-vel? Tegyük fel, hogy $X(t)$ Gauss-folyamat[5], azaz legyen $X(t) \sim N(0, \sigma_X^2)$. Mivel $a(\lambda_p)$ az $X(t)$-k lineáris kombinációja, így $a(\lambda_p)$ is normális eloszlású 0 várható értékkel, és

$$D^2 a(\lambda_p) = \frac{2\sigma_X^2}{N} \sum_{t=1}^{N} \cos^2 \lambda_p t = \begin{cases} \sigma_X^2, & \textit{ha } p \neq 0,\ p \neq \frac{N}{2} \ (\textit{ha } N \textit{ ps.}) \\ 2\sigma_X^2, & \textit{ha } p = 0 \textit{ vagy } p = \frac{N}{2} \textit{ és } N \textit{ ps.} \end{cases}$$

Teljesen hasonlóan $b(\lambda_p)$ is normális 0 várható értékkel, és

$$D^2 b(\lambda_p) = \frac{2\sigma_X^2}{N} \sum_{t=1}^{N} \sin^2 \lambda_p t = \begin{cases} \sigma_X^2, & \textit{ha } p \neq 0,\ p \neq \frac{N}{2} \ (\textit{ha } N \textit{ ps.}) \\ 0, & \textit{ha } p = 0 \textit{ vagy } p \neq \frac{N}{2} \textit{ és } N \textit{ ps.} \end{cases}$$

[3]könnyű számolás adja, hogy ez megegyezik az előzővel

[4]Fourier-frekvenciák

[5]összes $A_i = 0$ esete, tehát csak az $\varepsilon(t)$ zajt hagyjuk meg, és arról is feltesszük, hogy Gauss, azaz fehér zaj

szórásnégyzettel. Továbbá, még mindig az ortogonalitási relációból

$$cov(a(\lambda_p), b(\lambda_q)) = \frac{2\sigma_X^2}{N} \cdot \sum_{t=1}^{N} \cos(\lambda_p t) \cdot \sin(\lambda_q t) = 0$$

minden p, q-ra. Ugyancsak igaz $cov(a(\lambda_p), a(\lambda_q)) = 0$ és $cov(b(\lambda_p), b(\lambda_q)) = 0 \quad p \neq q$ esetén.

Viszont $a(\lambda_p)$ és $b(\lambda_p)$ is X_t-k lineáris kombinációja, így együttes eloszlásuk is normális, ráadásul a kölcsönös korrelálatlanság miatt függetlenek is. Mivel $I_p = I_N(\lambda_p)$ két független, 0 várható értékű normális eloszlású valószínűségi változó négyzetösszege, ezért $const \cdot \chi^2$ eloszlású. Tehát

$$I_p = (a(\lambda_p))^2 + (b(\lambda_p))^2 \sim \begin{cases} \sigma_X^2 \cdot \chi_2^2, & ha\ p \neq 0, \frac{N}{2}\ (N\ ps.) \\ 2 \cdot \sigma_X^2 \cdot \chi_1^2, & ha\ p = 0\ v.\ \frac{N}{2}\ (N\ ps.) \end{cases}.$$ [6]

Mivel a χ_k^2 várható értéke és szórása rendre k illetve $2k$, ezért $EI_p = 2\sigma_X^2$ minden p-re, és $D^2 I_p = 4\sigma_X^4$, ha $p \neq 0, \frac{N}{2}$, és $D^2 I_p = 8\sigma_X^4$, ha $p = 0, \frac{N}{2}$.

Szeretnénk megkapni $I_N(\lambda)$-t abban az esetben, amikor $\lambda \neq p \cdot \frac{2\pi}{N}$. Ehhez hasznos a következő

5.2. Lemma. $I_N(\lambda) = 2 \cdot \sum_{\tau=-(N-1)}^{N-1} \hat{R}(\tau) \cdot \cos(\tau\lambda)$.

Bizonyítás.

$$I_N(\lambda) = \frac{2}{N} \left| \sum_{t=1}^{N} X_t \cdot e^{-i\lambda t} \right|^2 = \frac{2}{N} \left(\sum_{t=1}^{N} X_t \cdot e^{-i\lambda t} \right) \left(\sum_{s=1}^{N} X_s \cdot e^{i\lambda s} \right) = \frac{2}{N} \left(\sum_{t=1}^{N} \sum_{s=1}^{N} X_t \cdot X_s \cdot e^{i\lambda(s-t)} \right) =$$

a szinuszos tagok kiesnek a szimmetria miatt, így az egyenlőség a következőképpen folytatható:

$$= \frac{2}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} X_t \cdot X_s \cdot \cos((s-t)\lambda) = (*)$$

A diagonálison összegezve $\tau = s - t$ mellett

$$(*) = 2 \cdot \sum_{\tau=-(N-1)}^{N-1} \left\{ \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t \cdot X_{t+|\tau|} \right\} \cdot \cos(\tau \cdot \lambda) = 2 \cdot \sum_{\tau=-(N-1)}^{N-1} \hat{R}(\tau) \cdot \cos(\tau \cdot \lambda),$$

amivel beláttuk a lemmát. Ezt persze a következőképpen is felírhatjuk:

$$2 \cdot \left(\hat{R}(0) + 2 \cdot \sum_{\tau=1}^{N-1} \hat{R}(\tau) \cdot \cos(\tau \cdot \lambda) \right).$$

■

Ez alapján kiszámolható $I_N(\lambda)$ várható értéke is, miszerint

$$EI_N(\lambda) = 2 \cdot \sum_{\tau=-(N-1)}^{N-1} E\hat{R}(\tau) \cdot \cos(\tau\lambda) = 2 \cdot \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N} \right) R(\tau) \cdot \cos(\tau\lambda).$$ [7]

Ha X_t-t $X_t = Z_t + \varepsilon_t$ alakban írjuk fel, akkor Z_t és ε_t függetlensége miatt $R(\tau) = R_Z(\tau) + R_\varepsilon(\tau)$ adódik, ahol $R_\varepsilon(\tau)$ a fehér zaj tulajdonságaiból adódóan σ_ε^2, ha $\tau = 0$, egyébként meg 0, továbbá

[6] megjegyzendő, hogy $\chi_2^2 \sim \left(\frac{1}{2}\right) - \exp$, így a periodogram a végpontokon kívül exponenciális eloszlású

[7] itt $R(\tau)$ az elméleti kovariancia, azaz $cov(X_t, X_{t+\tau})$

$R_Z(\tau) = \frac{1}{2} \cdot \sum_{i=1}^{K} A_i^2 \cdot \cos(\lambda_i \tau)$ (hiszen Z harmonikus folyamat). Ezt behelyettesítve (és a $\sum$-kat máris felcserélve)

$$EI_N(\lambda) = \sum_{i=1}^{K} A_i^2 \cdot \Big\{ \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) \cdot \cos(\tau\lambda_i) \cdot \cos(\tau\lambda) \Big\} + 2\sigma_\varepsilon^2.$$

Használjuk fel, hogy $\cos(\tau\lambda_i) \cdot \cos(\tau\lambda) = \frac{1}{2}\left(\cos(\tau(\lambda + \lambda_i)) + \cos(\tau(\lambda - \lambda_i))\right)$, illetve vegyük észre, hogy megjelent a Fejér-mag: $F_N(\vartheta) = \frac{1}{2\pi N} \cdot \frac{\sin^2\left(\frac{N\vartheta}{2}\right)}{\sin^2\left(\frac{\vartheta}{2}\right)} = \frac{1}{2\pi} \cdot \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) \cdot \cos(\tau\vartheta)$. Ezekből kapjuk, hogy

$$EI_N(\lambda) = 2\sigma_\varepsilon^2 + \frac{1}{2N} \cdot \sum_{i=1}^{K} A_i^2 \left(\frac{\sin^2\left(\frac{N(\lambda+\lambda_i)}{2}\right)}{\sin^2\left(\frac{\lambda+\lambda_i}{2}\right)} + \frac{\sin^2\left(\frac{N(\lambda-\lambda_i)}{2}\right)}{\sin^2\left(\frac{\lambda-\lambda_i}{2}\right)} \right) =$$

$$= 2\sigma_\varepsilon^2 + \pi \sum_{i=1}^{K} A_i^2 \cdot \big\{ F_N(\lambda + \lambda_i) + F_N(\lambda - \lambda_i) \big\}.$$

Az $F_N(\vartheta)$ Fejér-féle magfüggvénynek a $\vartheta = 0$-ban $\frac{N}{2\pi}$ nagyságrendű csúcsa van, gyorsan csökken, és $\frac{2k\pi}{N}$-ben 0-t vesz fel, $\frac{(2k+1)\pi}{N}$ környékén pedig kis, $O\left(\frac{1}{N}\right)$ nagyságrendű lokális maximumai vannak [ÁBRA]. Ha tehát a λ_i értékek elegendően távol vannak egymástól, azaz $|\lambda_i - \lambda_j| \gg \frac{2\pi}{N}$ minden $i \neq j$-re, akkor a periodogram a $\pm\lambda_i$ pontokra koncentrált Fejér-magok sorozata lesz [ÁBRA].

Azonban a periodogramot csak a $\lambda_p = 2\pi \cdot \frac{p}{N}$ pontokban számítjuk ki. Ha az igazi λ_i pont egybeesik valamely λ_p-vel, akkor $EI_N(\lambda_p)$-nek $N \cdot \frac{A_i^2}{2}$ nagyságú csúcsa lesz, ellenkező esetben a csúcs jelentősen megrövidülhet. A legrosszabb eset, ha a λ_i pont a λ_p és λ_{p+1} közötti felezőpontra esik. Whittle 1952-ben megmutatta, hogy ekkor a csúcs $\frac{4}{\pi^2}$-szeresére csökkenhet, így A_i^2 nagyságától függően akár el is veszíthetjük ezt a komponenst a vizsgálatok során ("leaking"). Ennek elkerülése végett a felosztást kiegészíthetjük még osztópontokkal[8], de ezek nem biztos, hogy $p \cdot \frac{2\pi}{N}$ alakúak, így elveszítjük a különböző frekvenciákon vett értékek függetlenségét, így a megfelelő tesztek érvényüket vesztik[9].

A szórás és a kovariancia kicsit több, de hasonló számolással kapható meg. Ettől most eltekintünk.

5.3. Tétel (Bartlett, 1955.).

$$cov\left(I_N(\lambda_1), I_N(\lambda_2)\right) = \frac{4\kappa_4}{N} + \frac{4\sigma_X^4}{N} \cdot 2\pi \cdot \left(F_N(\lambda_1 + \lambda_2) + F_N(\lambda_1 - \lambda_2)\right)$$

Ez alapján - mivel $F_N(\vartheta)$ a 0-ban $\frac{N}{2\pi}$ nagyságrendű, míg 0-tól távol $O\left(\frac{1}{N}\right)$ -

$$D^2\left(I_N(\lambda)\right) = \begin{cases} \frac{4\kappa_4}{N} + 4\sigma_X^4 + O\left(\frac{1}{N^2}\right), & \lambda \neq 0, \pm\pi \\ \frac{4\kappa_4}{N} + 8\sigma_X^4, & \lambda = 0, \pm\pi \end{cases}.$$

Azaz a Gauss folyamat esetén a λ_p alappontokban kapott eredmény $O\left(\frac{1}{N}\right)$ erejéig érvényben marad[10]. A kovarianciára pedig

$$cov\left(I_N(\lambda_1), I_N(\lambda_2)\right) = \begin{cases} 0, & ha\ X_t\ Gauss-folyamat,\ és\ \lambda_1, \lambda_2\ egyaránt\ \frac{2\pi}{N}\ többszörösei \\ O(\frac{1}{N^2}), & ha\ X_t\ Gauss-folyamat\ és\ |\lambda_1 \pm \lambda_2| \gg \frac{2\pi}{N} \\ O(\frac{1}{N}), & ha\ X_t-re\ \kappa_4 \neq 0\ és\ |\lambda_1 \pm \lambda_2| \gg \frac{2\pi}{N}\ vagy\ \lambda_1, \lambda_2\ \frac{2\pi}{N}\ többszörösei \end{cases}$$

tehát az aszimptotikus korrelálatlanság még nem Gauss esetben is fennáll, az alappontokon kívül is.

[8] "padding": 0-kal egészítik ki a sorozatot
[9] következő fejezet próbái
[10] 5.2 előtti számolás

A periodogram csúcsaira vonatkozó próbák

A standard frekvenciákon $\lambda_p = \frac{2\pi p}{N}$ ($p = 0, 1, \ldots, \left[\frac{N}{2}\right]$), a H_0 hipotézis pedig az, hogy valamennyi $A_i = 0$, vagyis $X(t)$ Gauss fehér zaj. Erre az esetre határoztuk meg $I_N(\lambda_p)$ (röviden I_p) eloszlását, amely $\frac{I_p}{\sigma_X^2} \sim \chi_2^2 \sim Exp(\frac{1}{2})$. Tekintsük most a periodogram legnagyobb értékét, $\max I_p$-t. Mivel I_p-k függetlenek, ez $\left[\frac{N}{2}\right]$ db független $Exp(\frac{1}{2})$ maximuma, ha leosztjuk még σ_X^2-tel, akkor

$$\gamma = \frac{\max\limits_{1 \le p \le \left[\frac{N}{2}\right]} (I_p)}{\sigma_X^2}$$

Ekkor γ eloszlása ismert σ_X^2 esetén: $P(\gamma \geq z) = 1 - P(\gamma < z) \overset{11}{=} 1 - \left(1 - e^{-\frac{z}{2}}\right)^{\left[\frac{N}{2}\right]}$. Ha $A_i \neq 0$ lenne valamely i-re, akkor γ "nagy" lenne, és egyoldali tesztet használhatnánk $\{\gamma > z_\alpha\}$ kritikus tartománnyal, z_α kritikus értéket az eloszlásból meghatározva[12] α terjedelem mellett.

A probléma viszont az, hogy σ_X^2 nem ismert, ezért csak becsülhetjük. Ha H_0 igaz, akkor $EI_p = 2\sigma_X^2$ minden $p \geq 1$-re, ezért $E \sum\limits_{p=1}^{\left[\frac{N}{2}\right]} I_p = 2 \cdot \left[\frac{N}{2}\right] \sigma_X^2$, így $v = \frac{1}{2\left[\frac{N}{2}\right]} \sum\limits_{p=1}^{\left[\frac{N}{2}\right]} I_p$ a σ_X^2-nek torzítatlan becslése. Ennek segítségével nagymintás tesztet konstruálhatunk $g^* = \frac{\max I_p}{v}$ próbastatisztikával. Nagy N-ekre a nevező fluktuálását figyelmen kívül hagyjuk, és $P(g^* \geq z) \sim 1 - \left(1 - e^{-\frac{z}{2}}\right)^{\left[\frac{N}{2}\right]}$ alapján készíthetünk egyoldali tesztet $\mathfrak{X} = \{g^* > z_\alpha\}$ szerint. Ez a Walker-féle teszt 1914-ből. Később, 1929-ben Fisher egzakt tesztet is meghatározott $g = \frac{\max I_p}{\sum\limits_{p=1}^{\left[\frac{N}{2}\right]} I_p}$ pontos eloszlása alapján, amelyet Fisher-féle g-próbának, más néven g-statisztikának nevezünk. Megmutatta, hogy H_0 mellett g eloszlása

$$P(g > z) = n \cdot (1-z)^{n-1} - \binom{n}{2}(1-2z)^{n-1} + \ldots + (-1)^a \binom{n}{a}(1-az)^{n-1},$$

ahol a az a legnagyobb egész szám, amely $\frac{1}{z}$-nél még kisebb. Ezek után ugyanúgy $P(g > z_\alpha) = \alpha$ alapján választunk z_α kritikus értéket. Ha a teszt elutasítja a nullhipotézist, akkor csak arra következtethetünk, hogy van **valamilyen** frekvenciájú periodikus komponens a folyamatban.[13] De vajon ez a frekvencia pont az a λ_p, amelyre I_p maximális? 1949-ben Hartley megmutatta, hogy annak a valószínűsége, hogy az I_p-beli legnagyobb csúcs valamilyen λ_p-től különböző λ frekvencia mellett következik be - ahol $|\lambda - \lambda_p| \gg \frac{2\pi}{N}$ -, kisebb, mint a teszt α szignifikanciaszintje. Tehát a csúcs helyét bátran elfogadhatjuk a frekvencia becslésének. 1952-ben Whittle még megmutatta, hogy ha λ_i-k megfelelően távol helyezkednek el egymástól, akkor $E\hat{\lambda}_i = \lambda_i + O(\frac{1}{N})$ és $D^2\hat{\lambda}_i = \frac{24\sigma_\varepsilon^2}{N^3(a_i^2+b_i^2)} + O\left(\frac{1}{N^3}\right)$. Ez utóbbi eredmény meglepő lehet, hiszen lineáris modelleknél a becslések szórásai általában $O\left(\frac{1}{N}\right)$ nagyságrendűek.

A Fisher-teszt csak az első, legnagyobb csúccsal foglalkozik, de Whittle javaslata (1952) szerint a tesztet ki lehet terjeszteni, kidobva a maximális tagot az I_p-k közül. Azaz legyen $g' = \frac{I^*_{p_{n-1}}}{\sum\limits_{p=1}^{\left[\frac{N}{2}\right]} I_p - I^*_{p_n}}$, és ennek az eloszlását a Fisher-féle g eloszlásból nyerjük úgy, hogy n-et $n-1$-re cseréljük.[14] Ez persze megtartja nullhipotézisnek a fehér zajt. Ha az így kapott g' is szignifikáns, megyünk tovább a következő legnagyobb csúcsra, és addig folytatjuk, amíg a tesztek szignifikáns eredményeket adnak. Így becslést kaphatunk a periodikus komponensek K számára. Az eljárás csak akkor működik, ha az igazi frekvenciák mind közel vannak $\frac{2\pi}{N}$ egész számú többszöröseihez. Egyébként az $\frac{A_k^2}{\sigma_\varepsilon^2}$ jel-zaj arány (signal to noise ratio) lesz lényeges, mert ha egy λ_k éppen két osztópont között van félúton, akkor a korábban már említett $\frac{4}{\pi^2}$ ($\sim 0,4$) -es redukció miatt ez a komponens detektálatlan maradhat[15]. Ekkor padding,

[11] a maximum pontosan akkor kisebb z-nél, ha mind kisebb (és persze I_p-k függetlenek)
[12] $P(\gamma > z_\alpha) = \alpha$-ból
[13] kiugró csúcsokra kell gondolni
[14] $I^*_{p_n}$ a rendezett minta n-edik eleme.
[15] elég nagy "signal to noise ratio" mellett a próba továbbra is alkalmazható

azaz 0-kal kiegészítés lehetséges, de ekkor a Fisher-teszt már nem használható.

Grenander és Rosenblatt meghatározták a

$$g_r = \frac{I^*_{p_{n-r}}}{\sum\limits_{p=1}^{\left[\frac{N}{2}\right]} I_p}$$

statisztikák eloszlását. Az ezzel a statisztikával konstruált próba csak akkor működik, ha a nullhipotézis hamissága esetén, a folyamat **pontosan** r periodikus komponenst tartalmaz. Sajnos ez szekvenciális tesztre nem alkalmas[16], mert amint találunk egy periodikus komponenst, a H_0 melletti eloszlás azonnal megváltozik, és így az újabb próbákra már nem használható.

5.4. Megjegyzés. *Ha több periodikus komponens van jelen közel ugyanakkora amplitúdóval, mint a maximális, akkor a Fisher-féle g-statisztika ereje csökken. Ugyanis* $\sum\limits_{p=1}^{\left[\frac{N}{2}\right]} I_p$ *tartalmazni fog néhány,* $\max I_p$*-vel azonos nagyságrendű tagot, így az értéke jóval nagyobb, ezért* g *értéke jóval kisebb lesz, mint egy periodikus komponens esetén.*

[16]nem tudjuk tesztek sorozatát alkalmazni, mint az előbb