

Az idősurelmélet alapfogalmai

Definíció. Sztochasztikus folyamat: $(X_t)_{t \in T}$, ahol T a paraméterter és minden t -re X_t valószínűségi változó.

Ebben a tárgyban általában $T = \mathbb{Z}$ vagy $T = \mathbb{Z}_+$, sztochasztikus folyamatokban $T = \mathbb{R}$ vagy $T = \mathbb{R}_+$.

Definíció. Gauss-folyamat: olyan sztochasztikus folyamat, melynek bármely véges számú peremeloszlása együttesen normális eloszlású, azaz minden $n \in \mathbb{Z}_+$, $t_1 \in T, \dots, t_n \in T$ esetén $(X_{t_1}, \dots, X_{t_n})$ együttesen normális eloszlású.

Idősor: Olyan sztochasztikus folyamat, amikor a T paramétertartományra 'idő'-ként gondolunk.

Definíció. Erős stacionaritás. $(X_t)_{t \in T}$ erősen stacionárius, ha minden $n \in \mathbb{Z}_+$, $t_1 \in T, \dots, t_n \in T$ és $h \in T$ esetén $(X_{t_1}, \dots, X_{t_n})$ együttesen ugyanolyan eloszlású, mint a h -val való eltoltja, $(X_{t_1+h}, \dots, X_{t_n+h})$.

Definíció. Autokovariancia függvény: $R(t, s) = \text{cov}(X_t, X_s)$

Definíció. Gyenge stacionaritás. $(X_t)_{t \in T}$ gyengén stacionárius, ha EX_t nem függ t -től (azaz konstans), illetve az autokovariancia függvény $R(t, s)$ értéke csak a $t - s$ eltéréstől függ.

Gyengén stacionárius idősor autokovariancia függvénye tehát tulajdonképpen egyváltozós, ezt az egyváltozós függvényt is R -rel fogjuk jelölni: $R(t, s) = R(t - s)$. Tehát gyengén stacionárius idősor autokovariancia függvénye $R(h) = \text{cov}(X_{t+h}, X_t)$ módon számolható.

Megjegyzés. A gyenge stacionaritásból nem következik az erős, de az erős stacionaritásból se a gyenge (nem biztos, hogy léteznek momentumai).

Megjegyzés. A *stacionaritás* szó bizonyos szempontból időbeli állandóságot, stabilitást jelent.

Megjegyzés. A stacionaritás közkedvelt, gyakori feltételezés egy tapasztalati idősorra vonatkozóan, azonban ellenőrzése nem egyszerű. Ha egy idősorban szemmel láthatóan trend vagy szezonális figyelhető meg, akkor nem stacionárius. Az idősorelemzés első lépése mindig a nemstacionárius összetevők (komponensek) kiszűrése.

A továbbiakban feltesszük, hogy az idősor gyengén stacionárius és paramétertere diszkrét.

Állítás. $R(0) = D^2 X_t$ minden t -re.

Definíció. Autokorreláció függvény (ACF): $r(h) = \text{cor}(X_t, X_{t+h})$, $h \in \mathbb{Z}$.

Állítás. $r(h) = \frac{R(h)}{R(0)}$

Definíció. Parciális autokorreláció függvény (PACF):

$\rho(h) = \text{cor}(X_t, X_{t+h} | X_{t+1}, X_{t+2}, \dots, X_{t+h-1})$, $h \in \mathbb{Z}$.

Jelölés.

• $r_h := r(h)$

• $R_h := \begin{bmatrix} 1 & r_1 & r_2 & \dots & r_{h-1} \\ r_1 & 1 & r_1 & \dots & r_{h-2} \\ r_2 & r_1 & 1 & \dots & r_{h-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{h-1} & r_{h-2} & r_{h-3} & \dots & 1 \end{bmatrix}$

• $R_h^\#$: a fenti R_h mátrix utolsó sorát kicseréljük r_1, \dots, r_h sorra.

Állítás. $\rho(h) = \begin{cases} r_1 & \text{ha } |h| = 1 \\ \frac{\det R_h^\#}{\det R_h} & \text{ha } |h| > 1 \end{cases}$

Definíció. Független értékű zaj folyamat: $\varepsilon_t \sim i.i.d.(0, \sigma^2)$, ha $E\varepsilon_t = 0$, $D^2\varepsilon_t = \sigma^2$, valamint ε_t és ε_s minden $t \neq s$ esetén független egymástól.

Definíció. Fehér zaj folyamat (white noise):

$\varepsilon_t \sim WN(0, \sigma^2)$, ha $E\varepsilon_t = 0$, $D^2\varepsilon_t = \sigma^2$ és $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$, ha $t \neq s$.

Megjegyzés. Gyakran kényelmes feltenni a fehér zajról, hogy Gauss-folyamat, ilyenkor Gauss-féle fehér zajról beszélünk (GWN).

Definíció. Lineáris folyamat:

$X_t = \mu + \sum_{j=-\infty}^{\infty} \beta_j \varepsilon_{t-j}$, ahol $\mu \in \mathbb{R}$, $\varepsilon_t \sim WN(0, \sigma^2)$ és $\sum_{j=-\infty}^{\infty} |\beta_j| < \infty$.

Megjegyzés. A lineáris folyamat definíciójában lévő ε_t zaj folyamatot *innovációnak* vagy *meghajtó folyamatnak* is szokták nevezni. Ez az elnevezés más modellek esetén is használatos.

Állítás. Az X_t lineáris folyamat stacionárius.

Definíció. MA(∞) folyamat: Olyan lineáris folyamat, amely nem függ a zaj jövőbeli értékeitől, azaz $\beta_{-1} = \beta_{-2} = \dots = 0$.

Megjegyzés. MA=moving average, magyarul mozgóátlag.

Megjegyzés. Ha adott egy x_1, \dots, x_n tapasztalati minta, akkor az adatokban lévő nemstacionárius komponens eltüntetésére alkalmas simítási eljárás lehet (nem ez az egyetlen) a *mozgóátlagolás*. Például ha az adataink negyedévesek, akkor érdemes 4 lépésben átlagokat számítani: $\tilde{x}_t = \frac{x_t + x_{t+1} + x_{t+2} + x_{t+3}}{4}$, $t = 1, 2, \dots, n - 3$. A mozgóátlagolással az idősor rövidebb lesz, adatokat veszítünk.

Eredmények az autokorreláció becsléséről

Legyen X_1, \dots, X_n minta egy stacionárius folyamatból, melynek várható értéke μ , autokovariancia függvénye $R(h)$ és autokorreláció függvénye pedig $r(h)$. Tekintsük a következő becsléseket:

- μ becslése: $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$;
- $R(h)$ becslése: $\hat{R}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)$, ahol $|h| < n$ egész;
- $r(h)$ becslése: $\hat{r}(h) = \frac{\hat{R}(h)}{\hat{R}(0)}$, ahol $h = -(n-1), \dots, -1, 0, \dots, n-1$.

Megjegyzés. Az autokovariancia becslésére több statisztika is használható, lásd az előadásjegyzet 4. részét, amely azok statisztikai tulajdonságait is tartalmazza.

Vezessük be a következő jelöléseket:

- $\mathbf{r}(h) = (r(1), \dots, r(h))^T$
- $\hat{\mathbf{r}}(h) = (\hat{r}(1), \dots, \hat{r}(h))^T$
- $\mathbf{0}_h$: h hosszú vektor 0 elemekkel
- \mathbf{I}_h : $h \times h$ méretű egységmátrix

Tétel. Bartlett-tétel. Legyen X_t lineáris folyamat, $\sum_{j=-\infty}^{\infty} \beta_j^2 |j| < \infty$. Ekkor

$(\frac{1}{n} \mathbf{W})^{-\frac{1}{2}} (\hat{\mathbf{r}}(h) - \mathbf{r}(h)) \xrightarrow[n \rightarrow \infty]{d} N_h(\mathbf{0}_h, \mathbf{I}_h)$, ahol \mathbf{W} az ún. Bartlett-féle mátrix, melynek elemei: $[\mathbf{W}]_{i,j} = \sum_{k=-\infty}^{\infty} [r(k+i)r(k+j) + r(k-i)r(k+j) + 2r(i)r(j)r^2(k) - 2r(i)r(k)r(k+j) - 2r(j)r(k)r(k+i)]$.

Megjegyzés. Az előző tétel állítását kissé pongyolán, de könnyebben érthetően úgy is írhatjuk, hogy $\hat{\mathbf{r}}(h) \approx N_h(\mathbf{r}(h), \frac{1}{n} \mathbf{W})$.

Következmény. Amennyiben X_1, \dots, X_n egy fehér zaj folyamatból származó minta, akkor $\hat{\mathbf{r}}(h) \approx N_h(\mathbf{0}_h, \frac{1}{n})$, ezáltal közös $1 - \alpha$ megbízhatóságú konfidenciaintervallum készíthető az egyes autokorrelációkra: $0 \pm \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}$.

Megjegyzés. Ha az előző következményben $\alpha = 0.05$, akkor közelítőleg a $[-\frac{2}{\sqrt{n}}; \frac{2}{\sqrt{n}}]$ intervallumbecslés adódik. Amikor \mathbf{R} segítségével egy idősor autokorreláció függvényét elkészítjük az *acf* függvény segítségével, akkor a két szaggatott kék vonal a $\pm \frac{2}{\sqrt{n}}$ értéknek felel meg. Most pedig következnek az a próba, amihez az eddigi előkészítés szükséges volt – legyen α az elsőfajú hiba valószínűségének előre rögzített értéke:

H_0 : a minta autokorrelálatlan folyamatból származik

H_1 : legalább az egyik autokorreláció nem 0

Próbastatisztika: $\mathbf{T}(\mathbf{X}) = \hat{\mathbf{r}}(n-1)$, ami egy $n-1$ elemű vektor

Elfogadási tartomány: $\mathcal{X}_e = \left\{ \mathbf{X} : \text{minden } 1 \leq i \leq n-1 \text{-re } |\hat{r}_i| \leq \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}} \right\}$

Tehát ha valamelyik tapasztalati autokorrelációra azt találjuk, hogy annak abszolút értéke meghaladja a $\frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}$ értéket, akkor a minta $100 \cdot (1-\alpha)\%$ -os megbízhatósággal nem autokorrelálatlan folyamatból származik. Ennél rendszerint többet is látunk, a belső "sáv"-on kívül eső autokorrelációk arra is utalnak, milyen jellegű nemstacionaritással szembesülünk, illetve milyen idősor modellel érdemes megpróbálni magyarázni az adatainkat.

Idősorok "illeszkedésvizsgálata"

A modellező fő célja az, hogy egy adott minta (tapasztalati idősor) esetén kiválassza azt a modellt, amelyik legjobban illeszkedik az adataira. Ez gyakran meglehetősen nehéz feladat. Az idősor modellek túlnyomó többsége valamilyen fehér zaj folyamatból (amit innovációnak is szokás hívni, és gyakran megfigyelési, mérési 'hiba'-ként vagy egyéb véletlen hatások összességéként tekintenek rá) indulnak ki. Az adott idősor modell illeszkedését úgy szokás megvizsgálni, hogy az illesztés – paraméterek becslése – után visszabecsüljük az innovációkat, majd megnézzük, hogy ezek vajon fehér zaj folyamatot követnek-e. Amennyiben a visszabecsült innovációk, más néven reziduálisok esetén *nem* vehető el, hogy fehér zaj folyamatból származnak, akkor az adott idősor modell alkalmazása ellen *nem* találtunk statisztikai bizonyítékot.

Jellemzően mik tudják elrontani a 'fehér zaj'-ságot:

- autokorreláció – a reziduálisok nem autokorrelálatlanok
- heteroszkedaszticitás – a reziduálisok szórása időben változik (\longleftrightarrow homoszkedaszticitás: a szórás időben konstans)

Próbák arra vonatkozóan, hogy egy adott X_1, \dots, X_n minta független fehér zajból származik-e:

a.) A minta autokorreláció függvény *pontonként* egyenlő-e 0-val. A próba leírásáért lásd az előző fejezetet.

Amikor \mathbf{R} segítségével egy idősor autokorreláció függvényét elkészítjük az *acf* függvény segítségével, akkor a két szaggatott kék vonal a $\pm \frac{2}{\sqrt{n}}$ értéknek felel meg. Az alábbi esetekben döntünk amellet, hogy a minta NEM fehér zajból származik:

- az egyik autokorreláció nagyon kiugrik, "sokkal" a kék sávon kívül van. Mi az a "sok": az adott autokorreláció abszolút értéke legalább akkora, mint a sáv hossza.
- ugyan nincs durván kiugró autokorreláció érték, de viszonylag sok érték a kék sávon kívül. Hüvelyujjszabály arra, mi minősül viszonylag soknak: ha m darab autokorrelációt rajzoltunk ki, akkor több, mint $0,05 \cdot m$ autokorreláció van kicsit a kék sávon kívül. Az R alap beállítása az $m = 40$, ezen a legalább 3 érték van kicsit a kék sávon kívül, akkor elvetjük

Ennél rendszerint többet is látunk, a belső "sáv"-on kívül eső autokorrelációk arra is utalnak, milyen jellegű nemstacionaritással szembesülünk, illetve milyen idősor modellel érdemes megpróbálni magyarázni az adatainkat.

b.) Portmanteau-tesztek. Nullhipotézisük: az első néhány (általában 20/30) autokorreláció *együttesen* egyenlő-e 0-val. Azaz ha H_0 nem teljesül, akkor NEM hisszük el, hogy a minta fehér zajból származik. A legfontosabbak ezek közül *Ljung-Box próba*. Próbastatisztika: $Q_{LB} = n(n+2) \sum_{i=1}^h \frac{\hat{r}(i)}{n-i}$, ami H_0 esetén χ_h^2 -hoz tart eloszlásban.

c.) *Különbség-előjel próba* (difference-sign test). Legyen Z_n annak a száma, ahányszor a folyamat értéke növekszik, azaz $Z_n = \sum_{i=2}^n Y_i$, ahol $Y_i = I(X_i > X_{i-1})$, $i = 2, 3, \dots, n$. A próbastatisztika: $\frac{Z_n - \frac{n-1}{2}}{\sqrt{\frac{n-1}{12}}}$, ami standard normális eloszláshoz tart eloszlásban

a nullhipotézis esetén.

d.) *Fordulópont próba* (turning point test). Legyen Z_n a fordulópontok száma a mintában. Az x_i mintapont fordulópont, amennyiben vagy $(x_i > x_{i-1} \text{ és } x_i > x_{i+1})$, vagy $(x_i < x_{i-1} \text{ és } x_i < x_{i+1})$. Megmutatható, hogy a $\frac{Z_n - \frac{2}{3}(n-2)}{\sqrt{(n-2)\frac{8}{45}}}$ próbastatisztika standard normális eloszlásban tart eloszlásban a nullhipotézis esetén.

Heteroskedaszticitás próba idősorok esetén: *Mcleod Li teszt*.

Definíció. m -összefüggőség. Az $(X_n)_{n \in \mathbb{Z}}$ valószínűségi változó sorozat m -összefüggő, amennyiben a $\sigma(\dots, X_{k-1}, X_k)$ és $\sigma(X_{k+m+1}, X_{k+m+2}, \dots)$ σ -algebrák minden $k \in \mathbb{Z}$ -re függetlenek.

Megjegyzés. Az egységgyök nélküli MA(p) folyamat p -összefüggő.

Tétel. Centrális határeloszlás-tétel m -összefüggőség esetén.

Legyen X_1, X_2, \dots , m -összefüggő, gyengén stacionárius valószínűségi változó sorozat,

$$\sigma_m^2 := D^2 X_1 + 2 \sum_{k=1}^m \text{cov}(X_1, X_{1+k}). \quad \text{Ekkor } \frac{\sum_{i=1}^n (X_i - EX_i)}{\sqrt{n} \sigma_m} \xrightarrow[n \rightarrow \infty]{d} N(0; 1).$$

ARMA folyamatok

Definíció. Visszaléptetés operátor (backshift operator).

$$B : L_2(\Omega) \rightarrow L_2(\Omega), \quad BX_t = X_{t-1}$$

Állítás. A fenti B operátor

- iterálható: $B^k X_t = X_{t-k}$, $k \in \mathbb{Z}_+$;
- invertálható: $B^{-1} X_t = X_{t+1}$;
- unitér.

Definíció. ARMA folyamat. $X_t = \underbrace{\sum_{i=1}^p \alpha_i X_{t-i}}_{\text{AR}(p) \text{ rész}} + \underbrace{\sum_{j=0}^q \beta_j \varepsilon_{t-j}}_{\text{MA}(q) \text{ rész}}$, ahol p, q pozitív egész

(nevük: *rend*); $\alpha_i, \beta_j \geq 0$ valós számok; $\varepsilon_t \sim WN(0, \sigma^2)$.

Megjegyzés. ARMA: autoregresszív mozgóátlag (autoregressive moving average)

Megjegyzés. Az ARMA folyamat másik alakja: $\sum_{i=0}^p \tilde{\alpha}_i X_{t-i} = \sum_{j=0}^q \beta_j \varepsilon_{t-j}$, ahol $\tilde{\alpha}_0 = 1$ és $\tilde{\alpha}_i = -\alpha_i$ ($i = 1, \dots, p$).

Definíció. Kauzalitás. Az ARMA(p, q) folyamat kauzális (vagy okozati), ha léteznek olyan ψ_i , $i \in \mathbb{Z}$ konstansok, amikre $\sum_{i=1}^{\infty} |\psi_i| < \infty$ és $X_t = \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i} \forall t$ -re.

Megjegyzés. Az okozatiság a következőket jelenti:

- az ARMA "egyenletnek" létezik "megoldása";
- az ARMA folyamatnak létezik MA(∞) alakja.

Definíció. Invertálhatóság. Az ARMA(p, q) folyamat invertálható, ha létezik olyan π_i , $i \in \mathbb{Z}$ konstansok, amikre $\sum_{i=1}^{\infty} |\pi_i| < \infty$ és $\varepsilon_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} \forall t$ -re.

Megjegyzés. Az invertálhatóság helyett azt is szokás mondani, hogy a folyamatnak van AR(∞) alakja.

Definíció. ARMA folyamat karakterisztikus polinomjai: $P(x) = \sum_{i=0}^p \tilde{\alpha}_i x^{p-i}$,

$$\tilde{P}(x) = \sum_{i=0}^p \tilde{\alpha}_i x^i, \quad Q(x) = \sum_{i=0}^q \beta_i x^i.$$

Állítás. $P(x) = x^p \cdot \tilde{P}\left(\frac{1}{x}\right)$

Állítás. $P(x)$ gyökei az egységkörön belül vannak $\iff \tilde{P}(x)$ gyökei az egységkörön kívül vannak

Tétel. Az ARMA folyamat "megoldása" és "inverze".

- Ha $P(x)$ gyökei az egységkörön belül vannak, akkor a folyamatnak létezik stacionárius megoldása;
- Ha $Q(x)$ gyökei az egységkörön kívül vannak, akkor a folyamatnak létezik inverze.

Következmény. Az $X_t = \alpha X_{t-1} + \varepsilon_t$ AR(1) folyamat stacionárius $\iff |\alpha| < 1$

Tétel. Az ARMA folyamat spektrális sűrűségfüggvénye: $f(x) = \frac{\sigma^2}{2\pi} \left| \frac{Q(e^{ix})}{P(e^{ix})} \right|^2$

Eddig jelöléseinkkel az ARMA folyamat a következő operátoros alakba írható: $\tilde{P}(B)X_t = Q(B)\varepsilon_t$. Ezt már ki tudjuk fejezni az ARMA folyamatra, illetve a fehér zajra is:

- $X_t = [\tilde{P}(B)]^{-1} Q(B)\varepsilon_t$
- $\varepsilon_t = [Q(B)]^{-1} \tilde{P}(B)X_t$

Az persze nem triviális, hogy a fenti operátorinverzeket, majd az operátorszorzatokat hogyan állítsuk elő, ezen a ponton segítségül kell hívni a homogén lineáris differenciaegyenletek elméletét. Legyen célunk a fenti X_t előállítás explicit megadása, a másik hasonlóan megy. Jelölje $S(x) = \sum_{i=0}^{\infty} s_i x^i$ azt a polinomot, amire $S(B) = [\tilde{P}(B)]^{-1} Q(B)$.

Ebben az inverz úgy kapható meg, hogy keressük azt a $T(x) = \sum_{i=0}^{\infty} t_i x^i$ polinomot,

amire $T(x)\tilde{P}(x) = 1$. Ezen a ponton rendszerint abba futunk bele, hogy az s_i együttműködőkre egy rekurzió (differenciaegyenlet) adódik, amit meg kellene oldani. Tekintsük a következő homogén differenciaegyenletet:

$$s_t + \gamma_1 s_{t-1} + \dots + \gamma_k s_{t-k} = 0,$$

ahol

- $k \leq t \in \mathbb{Z}$;
- $\gamma_1, \dots, \gamma_k \in \mathbb{Z}$; $\gamma_i \neq 0 \forall i$ -re;
- adottak s_0, s_1, \dots, s_{k-1} kezdeti értékek.

Mi a gyakorlaton "kézzel" legfeljebb másodrendű differenciaegyenleteket fogunk megoldani. Ezekről szólnak az alábbi állítások.

Állítás. Az $s_t + \gamma_1 s_{t-1} = (1 - \xi^{-1}B)s_t = 0$ elsőrendű homogén lineáris differenciaegyenlet általános megoldása: $s_t = s_0 \xi^{-t}$.

Állítás. Az $s_t + \gamma_1 s_{t-1} + \gamma_2 s_{t-2} = (1 - \xi_1^{-1}B)(1 - \xi_2^{-1}B)s_t = 0$ másodrendű homogén lineáris differenciaegyenlet általános megoldása:

1. eset: $\xi_1 \neq \xi_2$ valósak $\rightsquigarrow s_t = c_1 \xi_1^{-t} + c_2 \xi_2^{-t}$, ahol c_1 és c_2 valós konstansok a kezdeti értékek felhasználásával, a $\begin{cases} c_1 + c_2 & = s_0 \\ c_1 \xi_1^{-1} + c_2 \xi_2^{-1} & = s_1 \end{cases}$ egyenletrendszer megoldásából adódnak (a többinél hasonlóan).
2. eset: $\xi_1 = \xi_2$ valósak $\rightsquigarrow s_t = (c_1 + c_2 t) \xi_1^{-t}$, ahol c_1 és c_2 valós konstansok a kezdeti értékek felhasználásával adódnak.
3. eset: $\xi_1 \neq \xi_2$ komplexek $\rightsquigarrow s_t = c_1 \xi_1^{-t} + c_2 \xi_2^{-t}$, ahol c_1 és c_2 komplex konstansok a kezdeti értékek felhasználásával adódnak. Az általános megoldás másképp is felírható, mivel algebrából megtanultuk, hogy $\xi_2 = \bar{\xi}_1$. Felhasználva a $\xi_1 = de^{i\theta}$ exponenciális alakot, $s_t = ad^{-t} \cos(\theta t + b)$, ahol a és b a kezdeti értékek felhasználásával adódó valós konstansok.

Az ARMA modellek kiválasztásánál segítségünkre lehetnek azok tulajdonságai:

Modell	$R(h)$ autokorreláció fv.	$r(h)$ parciális autokorreláció fv.
AR(p)	tart 0-hoz, ha $ h \rightarrow \infty$	nem 0, ha $ h \leq p$; egyébként 0
MA(q)	nem 0, ha $ h \leq q$; egyébként 0	tart 0-hoz, ha $ h \rightarrow \infty$
ARMA(p, q)	tart 0-hoz, ha $ h \rightarrow \infty$, az első q érték után kezdődik a konvergencia	tart 0-hoz, ha $ h \rightarrow \infty$, az első p érték után kezdődik a konvergencia

Modellválasztás: a "legjobb" ARMA(p, q) folyamat kiválasztása. Lépései:

1. Rende (p, q) kiválasztása.

Ebben segítségünkre vannak az alábbi információs kritériumok:

- **Akaike-féle információs kritérium:** $AIC = 2k - 2 \log \hat{L}$, ahol
 - k : a becslendő paraméterek száma
 - \hat{L} a likelihood-függvény értéke akkor, ha az ML-beclést használjuk (normális eloszlású hibáknál ez megegyezik a legkisebb négyzetes becléssel)
Minél kisebb, annál jobb.
- **Bayes-féle információs kritérium:** $BIC = \log n \cdot k - 2 \log \hat{L} \rightsquigarrow$ minél kisebb, annál jobb

A tapasztalati autokovariancia függvény alapján a fenti táblázat segíthet a megfelelő rend kiválasztásában. Az információs kritériumokon alapuló döntés rendszerint jobb.

2. Ismeretlen paraméterek beclése.

A klasszikus módszerek most is használhatók: momemtum módszer (az autokovarianciákból adódó Yule-Walker egyenletrendszer), ML-módszer, legkisebb négyzetek módszere. Az ML-módszer általában a legjobb, de mivel idősoros

modelleknél optimalizálás rutinnal számolódik, ezért rendszerint a leglassabb is egyben.

Becslés után meg kell nézni, hogy az együtthatók szignifikánsak-e. Minden számítógépes programcsomag meg szokta adni a paraméterek $\hat{\vartheta}$ pontbecslését és ezek $\widehat{D}(\hat{\vartheta})$ standard hibáját. Hüvelykujjszabály arra, hogy a ϑ paraméter szignifikáns-e, azaz állíthatjuk-e, hogy eltér-e nullától: a 0 benne van-e a $\hat{\vartheta} \pm 2 \cdot \widehat{D}(\hat{\vartheta})$ intervallumban. Ha nincs benne a 0, akkor szignifikáns.

Modelldiagnosztika: miután illesztettünk egy modellt, meg kell nézni, teljesülnek-e az adott idősortmodell definíciójának feltételei. ARMA modellek esetén ez annak vizsgálatát jelenti, hogy a reziduálisok származhatnak-e fehér zajból. Az alábbi tesztek szokás rendszerint végrehajtani:

- Empirikus autokorreláció függvényben a sávon kívüli értékek vizsgálata
- Ljung-Box próba vagy egy általánosítása
- Heteroszkedaszticitás teszt – amennyiben azt kapjuk, hogy az idősor nem homoszkedasztikus, akkor a reziduálisokra GARCH modellt illesztünk

Megjegyzés. A modelldiagnosztikába néha bele szokták venni az együtthatók szignifikanciájának vizsgálatát is.

Amennyiben azt kapjuk, hogy az ARMA folyamat nem illeszkedik megfelelően, számos további teendőnk lehet:

- Van-e trend/szezonalitás az idősorban (ezek kiszűrésével illik kezdeni az elemzést, lásd a következő fejezetet)
- Próbálkozás másik idősortmoddellel

Végül kimondjuk az idősorelmélet egyik legfontosabb tételét, ami megmutatja, miért van kiemelt szerepe az MA(∞) modelleszaládnak. A tételbeli 'négyzetes előrejelezhetőség' és a 'determinisztikusság' a viszonylag hosszadalmas előkészítés igénye miatt nem lesz definiálva, ld. például a Brockwell–Davis könyv 2.6. fejezetét és az előtte lévő fejezete(ke)t további részletekért.

Tétel. Wold-felbontás. Tetszőleges X_t stacionárius folyamat felírható a következő alakba: $X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} + V_t$, ahol

- $\psi_0 = 1, \sum_{i=0}^{\infty} \psi_i^2 < \infty$;
- $\varepsilon_t \sim WN(0, \sigma^2)$;
- V_t 'determinisztikus' folyamat;
- $\text{cov}(\varepsilon_t, V_s) = 0 \quad \forall s, t$ -re;
- ε_t és V_t 'négyzetesen előrejelezhető' folyamatok.

GARCH folyamatok

Definíció. GARCH folyamat. $(X_t)_{t \in \mathbb{Z}}$ GARCH(p, q) folyamat, ha $X_t = \sigma_t \varepsilon_t$, ahol

$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$ a folyamat időben változó szórásnégyzete,
 $\varepsilon_t \sim i.i.d.(0; 1)$, $E\varepsilon_t^4 < \infty$, továbbá $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ minden i és j esetén.

Speciálisan, ha $\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2$, akkor az X_t folyamatot ARCH(p) folyamatnak hívjuk.

Megjegyzés. A folyamatot $X_t = \sqrt{\alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2} \cdot \varepsilon_t$ alakba is írhatjuk.

Megjegyzés. A folyamat elnevezése: GARCH = generalised autoregressive conditional heteroskedastic, azaz magyarul általánosított autoregresszív, feltételesen heteroszkedasztikus.

Állítás. A GARCH(p, q) folyamat (NEM független) fehér zaj.

Tétel. A GARCH folyamat gyenge stacionaritása.

- Ha $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$, akkor a GARCH(p, q) folyamatnak létezik gyengén stacionárius megoldása ("megoldás": X_t folyamat MA(∞) alakja).
- Ha a GARCH(p, q) folyamatnak létezik gyengén stacionárius megoldása és $\alpha_0 > 0$, akkor $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$.

A GARCH folyamatból generált minta jellemzői:

- Az adatok nem korreláltak, és a szórás változik az idővel;
- Az adatok eloszlása vastag szélű;
- A négyzetek és az abszolútértékek erősen korreláltak;
- A kiugró értékek klaszterekben jelennek meg.

Amennyiben egy tapasztalati mintánál a fenti jellemzőket figyeljük meg, érdemes megpróbálkozni GARCH folyamattal modellezni. Ezek a tulajdonságok a pénzügyi adatsoroknál gyakran megfigyelhetők (például részvényárfolyamok, valutaárfolyamok loghozamainál).

Megjegyzés. Amennyiben valamilyen heteroszkedasztikus folyamattal szeretnénk modellezni a megfigyeléseket, akkor nagyon gyakran megfelelő választás a viszonylag egyszerű GARCH(1,1) modellt illeszteni.

Idősorok előrejelzése

Előrejelzési feladat: legyen $(X_n)_{n \in \mathbb{Z}}$ stacionárius folyamat μ várható értékkel és $R(h)$ autokovariancia függvényvel. Célunk a folyamat első n értéke, X_n, \dots, X_1 segítségével az ezt követő h érték, X_{n+h}, \dots, X_{n+1} meghatározása úgy, hogy a legkisebb négyzetes értelemben vett hiba a lehető legkisebb legyen.

Jelölje a legjobb lineáris előrejelzést $\mathbb{P}_n X_{n+h} := a_0 + a_1 X_n + \dots + a_n X_1$.

Állítás. A legjobb lineáris előrejelzés operátor tulajdonságai.

a.) A legjobb lineáris előrejelzés együtthatói kielégítik a következő egyenleteket:

$$E(X_{n+h} - \mathbb{P}_n X_{n+h}) = 0$$

$$E[(X_{n+h} - \mathbb{P}_n X_{n+h})X_{n+1-j}] = 0 \quad j = 1, \dots, n$$

b.) a legjobb lineáris előrejelzés együtthatói kielégítik a következő egyenletrendszert:

$$\mathbf{\Gamma}_n \mathbf{a}_n = \mathbf{R}_n(h), \text{ ahol}$$

$$\bullet \mathbf{\Gamma}_n = [R(i-j)]_{i,j=1}^n;$$

$$\bullet \mathbf{a}_n = (a_1, \dots, a_n)^T;$$

$$\bullet \mathbf{R}_n(h) = (R(h), R(h+1), \dots, R(h+n-1))^T.$$

c.) Az előrejelzés $\mathbb{P}_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$ alakba írható, így a feladatok elején $\mu = 0$ feltehető, majd ezzel a képlettel megkapható az előrejelzés;

d.) Az előrejelzés legkisebb négyzetes hibája $R(0) - \mathbf{a}_n^T \mathbf{R}_n(h)$.

Bizonyítás. MSE=mean square error = legkisebb négyzetes hiba

$$MSE(a_0, a_1, \dots, a_n) = E(X_{n+h} - a_0 - a_1 X_n - \dots - a_n X_1)^2 \longrightarrow \min_{a_0, a_1, \dots, a_n}$$

a.) Keressük a minimumot deriválással, deriváljuk az MSE függvényt az a együtthatók szerint.

$$\partial_{a_0} MSE = (-2) \cdot E\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}\right) = 0 \quad \Rightarrow$$

$$\Rightarrow \mu - a_0 - \sum_{i=1}^n a_i \mu = 0 \quad \Rightarrow \quad a_0 = \mu \left(1 - \sum_{i=1}^n a_i\right)$$

$$j = 1, 2, \dots, n\text{-re } \partial_{a_0} MSE = (-2) \cdot E\left[X_{n+1-j} \left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}\right)\right] = 0$$

b.) Írjuk be a_0 -t!

$$\Rightarrow E\left[X_{n+1-j} \left((X_{n+h} - \mu) - \sum_{i=1}^n a_i (X_{n+1-i} - \mu)\right)\right] = 0$$

$$\Rightarrow E(X_{n+1-j}(X_{n+h} - \mu)) = \sum_{i=1}^n a_i E(X_{n+1-j}(X_{n+1-i} - \mu))$$

A baloldalt továbbfejtvé,

$$E(X_{n+1-j}(X_{n+h} - \mu)) = E(X_{n+1-j}(X_{n+h} - \mu)) - \underbrace{E(X_{n+1-j}) \cdot E(X_{n+h} - \mu)}_{=0} =$$

$$= \text{cov}(X_{n+1-j}, X_{n+h} - \mu) = \text{cov}(X_{n+1-j}, X_{n+h}) = R(h+j-1)$$

Hasonló levezéssel, a jobboldalon lévő várható értékre $R(i-j)$ adódik. Így az egyenletre a következőt kapjuk:

$$R(h+j-1) = \sum_{i=1}^n a_i R(i-j) = [a_1, \dots, a_n] \begin{bmatrix} R(1-j) \\ \vdots \\ R(n-j) \end{bmatrix} \quad j = 1, 2, \dots, n$$

Írjuk az egyenleteket mátrix alakba:

$$\begin{bmatrix} R(h) \\ R(h+1) \\ R(h+2) \\ \vdots \\ R(h+n-1) \end{bmatrix} = \begin{bmatrix} R(0) & R(-1) & R(-2) & \cdots & R(-(n-1)) \\ R(1) & R(0) & R(-1) & \cdots & R(-(n-2)) \\ R(2) & R(1) & R(0) & \cdots & R(-(n-3)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(n-1) & R(n-2) & R(n-3) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix},$$

ami tömörebben épp az állítást adja: $\mathbf{R}_n(h) = \mathbf{\Gamma}_n \mathbf{a}_n$.

c.) $\mathbb{P}_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1$, amibe írjuk be az a.) feladatban a_0 -ra kapott optimális értéket:

$$\mathbb{P}_n X_{n+h} = \mu \left(1 - \sum_{i=1}^n a_i \right) + \sum_{i=1}^n a_i X_{n+1-i} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$$

$$\begin{aligned} \text{d.) } MSE(a_0, \dots, a_n) &= E(X_{n+h} - \mathbb{P}_n X_{n+h})^2 \stackrel{\text{(c.)}}{=} D^2 \left(X_{n+h} - \mu - \sum_{i=1}^n a_i (X_{n+1-i} - \mu) \right) = \\ &= D^2 X_{n+h} - 2 \sum_{i=1}^n a_i \text{cov}(X_{n+h}, X_{n+1-i} - \mu) + D^2 \left(\sum_{i=1}^n a_i (X_{n+1-i} - \mu) \right) = \\ &= R(0) - 2 \sum_{i=1}^n a_i R(h+i-1) + \sum_{j=1}^n \sum_{i=1}^n a_i a_j R(i-j) = \\ &= R(0) - 2 \mathbf{a}_n^T \mathbf{R}_n(h) + \mathbf{a}_n^T \underbrace{\mathbf{\Gamma}_n(h) \mathbf{a}_n}_{\mathbf{R}_n(h)} = R(0) - \mathbf{a}_n^T \mathbf{R}_n(h) \quad \square \end{aligned}$$

Állítás. $R(0) > 0$ és $R(h) \xrightarrow{h \rightarrow \infty} 0$ esetén a fenti $\mathbf{\Gamma}_n$ invertálható, így az a_i együtthatók $\mathbf{a}_n = \mathbf{\Gamma}_n^{-1} \mathbf{R}_n(h)$ módon számolhatók.

Az általános előrejelzési operátor.

Legyen $\mathbf{W} = (W_n, \dots, W_1)^T$ véletlen vektor, Y valószínűségi változó, mindannyian véges szórással/szórás mátrixszal. Jelölje $\mathbf{\Gamma} = \text{cov}(\mathbf{W}) = \Sigma(\mathbf{W})$ a kovarianciamátrixot. Célunk: \mathbf{W} segítségével Y legkisebb négyzetes lineáris előrejelzésének (becslésének) előállítás.

Jelölje az előrejelzési operátort $\mathbb{P}(\bullet | \mathbf{W}) : L_2(\Omega) \rightarrow L_2(\Omega)$, amit lineáris alakban keresünk: $\mathbb{P}(Y | \mathbf{W}) = a_0 + \mathbf{a}_n^T \mathbf{W}$, ahol a_0 és $\mathbf{a}_n = (a_1, \dots, a_n)^T$ a becsülendő értékek.

Állítás. Az általános legjobb lineáris előrejelzés operátor tulajdonságai.

Legyenek $\alpha_1, \dots, \alpha_n, \beta$ tetszőleges valós számok és Z tetszőleges véges szórású valószínűségi változó. Ekkor

a.) $\mathbb{P}(Y | \mathbf{W}) = EY + \mathbf{a}_n^T (\mathbf{W} - E\mathbf{W})$, ahol $\mathbf{\Gamma} \mathbf{a}_n = \text{cov}(Y, \mathbf{W})$;

b.) $E(Y - \mathbb{P}(Y | \mathbf{W})) = 0$

$$E[(Y - \mathbb{P}(Y | \mathbf{W})) \mathbf{W}] = 0;$$

c.) $MSE = E[(Y - \mathbb{P}(Y | \mathbf{W}))^2] = D^2 Y - \mathbf{a}_n^T \text{cov}(Y, \mathbf{W})$;

d.) $\mathbb{P}(\alpha_1 Y + \alpha_2 Z + \beta | \mathbf{W}) = \alpha_1 \mathbb{P}(Y | \mathbf{W}) + \alpha_2 \mathbb{P}(Z | \mathbf{W}) + \beta$;

e.) $\mathbb{P} \left(\sum_{i=1}^n \alpha_i W_i + \beta \middle| \mathbf{W} \right) = \sum_{i=1}^n \alpha_i W_i + \beta$;

f.) ha $\text{cov}(Y, \mathbf{W}) = \mathbf{0}$, akkor $\mathbb{P}(Y | \mathbf{W}) = EY$.

Megjegyzés. \mathbb{P} általános előrejelzési operátor kapcsolata a \mathbb{P}_n operátorral: $\mathbb{P}_n(X_{n+h}) = \mathbb{P}(X_{n+h} | \mathbf{X}_n)$, ahol $\mathbf{X}_n = (X_n, \dots, X_1)^T$.

Megjegyzés. A fenti állítás a.) részéből következik, hogy \mathbf{a}_n^T ismeretében már meghatározható az a_0 konstans: $a_0 = EY - \mathbf{a}_n^T E\mathbf{W}$. Speciális esetben, amennyiben a \mathbb{P}_n operátorral dolgozunk, akkor $a_0 = \mu \cdot \left(1 - \sum_{i=1}^n a_i \right)$.

Megjegyzés. A fenti állítás b.) részéből látszik, hogy a legjobb lineáris legkisebb négyzetes becslés torzítatlan.

Megjegyzés. Az általános előrejelzési operátor segítségével módunkban áll hiányzó vagy hibásnak vélt adatok értékét megbecsülni.

Nemstacionárius idősorok modellezése

Definíció. Lag-1 differencia operátor. $\nabla = 1 - B$, ahol B a visszaléptetés operátor.

Definíció. Lag-d differencia operátor. $\nabla_d = 1 - B^d$.

Ezáltal $\nabla X_t = X_t - X_{t-1}$ és $\nabla_d X_t = X_t - X_{t-d}$. A differencia operátort tetszőleges pozitív egész hatványra lehet emelni, ekkor $\nabla^m X_t = \nabla^{m-1}(X_t - X_{t-1})$, ami tovább iterálható.

Definíció. ARIMA modell. Az X_t folyamat ARIMA(p, d, q) folyamatot követ, amennyiben $Y_t = (1 - B)^d X_t$ folyamat ARMA(p, q) folyamatból származik.

Megjegyzés. Az ARIMA-ban az I betű az 'Integrated' angol szó rövidítése (integrált).

ARIMA modelleknél az integráltságot kifejező d paraméter értékének megállapításában az ún. **egységgyök tesztek** segítenek. Amennyiben a modell $P(x)$ karakterisztikus polinomjának van egységgyöke (az egyik gyökének 1 az abszolútértéke), akkor a folyamat nem lesz stacionárius. Az egységgyök tesztek olyan nem-stacionaritást vizsgálnak, amely differencia-képzéssel eltüntethető.

A leggyakrabban használt egységgyök teszt a **kiterjesztett Dickey-Fuller próba**. Működésének megértéséhez tekintsük először legegyszerűbb változatát.

Legyen $X_t = \alpha X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim GWN(0, \sigma^2)$ AR(1) folyamat. Tudjuk, hogy $\alpha = 1$ esetén ez éppen a véletlen bolyongás, ami nem stacionárius és $X_t = \sum_{i=1}^t \varepsilon_i$ alakba írható;

$|\alpha| < 1$ esetén pedig stacionárius. Az AR(1) egyenlet átírható a $\nabla X_t = \phi X_{t-1} + \varepsilon_t$ alakba, ahol $\phi = \alpha - 1$. A teszt hipotézisei: $H_0 : \alpha = 1$ és $H_1 : |\alpha| < 1$, tehát az ellenhipotézis azt állítja, hogy stacionárius a folyamat; a nullhipotézist pedig azt, hogy ugyan NEM stacionárius, de differencia képzésével stacionáriussá tehető. Végrehajtásához írjuk fel ∇X_t lineáris regresszióját X_{t-1} segítségével, a próbastatisztika pedig legyen a regressziós együttható standardizáltja: $\frac{\hat{\alpha}-1}{D(\hat{\alpha}-1)}$. A Donsker-tétel segítségével meg lehet mutatni, hogy a nullhipotézis esetén ez a próbastatisztika eloszlásban egy olyan valószínűségi változóhoz tart, aminek kvantiliseit csak szimulációval vagy numerikus approximációval lehet megkapni.

A próba kiterjesztett változata AR(p) folyamatból indul ki, valamint megengedi μ drift és ηt trend jelenlétét is a folyamatban: $X_t = \mu + \eta t + \sum_{j=1}^p \alpha_j X_{t-j} + \varepsilon_t$. Itt a regresszió

óra alkalmas alak $\nabla X_t = \mu + \eta t + \phi X_{t-1} + \sum_{j=1}^p \psi_j \nabla X_{t-j} + \varepsilon_t$, ahol $\phi = \sum_{j=1}^p \alpha_j - 1$ és

$\psi_j = - \sum_{k=j}^p \alpha_k$. A nem-stacionaritást most is $H_0 : \phi = 0$ -val teszteljük, a próbastatisztika itt is a ϕ alkalmas becslése, felhasználva a regressziós együtthatók becslését.

Egységgyök teszt végrehajtásánál a modellező feladata annak meghatározása, hogy mi legyen a p rend, valamint hogy drift és/vagy trend jelenlétére is számít-e a vizsgált folyamatnál.

Klasszikus idősor-dekompozíciós modell: $X_t = m_t + s_t + Y_t$, ahol

- m_t : trend komponens – valamilyen szabályosan változó függvény, ami gyakran lineáris vagy négyzetes.
- s_t : szezonális komponens – a rendszeresen ismétlődő, azonos periodicitású és szabályos amplitúdójú, rendszerint rövid távú ingadozásokat tartalmazza. Ha $d \in \mathbb{Z}$ jelöli a szezonalitást leíró periódusok hosszát, akkor $s_t = s_{t+d}$ minden t -re. Feltesszük, hogy $\sum_{i=1}^d s_i = 0$. Közgazdasági alkalmazásokban a d értéke negyedéves adatoknál jellemzően 4, havi adatoknál pedig 12.
- Y_t : véletlen zaj tag, egy olyan stacionárius komponens, amit már valamilyen ismert idősor-moddal modellezhetünk. Feltesszük, hogy $EY_t = 0$, különben a konstans beolvasztható lenne a trend tagba.

A nemstacionárius idősoroknál a trend hatás kimutatására, illetve eltüntetésére két megközelítést lehet követni:

1. Trend becslése

- *Paraméteres:* Alkalmos függvényt illesztünk, ami rendszerint egy polinom szokott lenni, azaz $m_t = \sum_{i=0}^p c_k t^k$ alakú, a c_k együtthatókat pedig legkisebb négyzetek módszerével lehet becsülni. Ezen a ponton kihasználhatjuk, hogy egy ilyen alakú regressziós modell a lineáris modell speciális esete. A paraméteres megközelítés hátránya, hogy feltesszük, a választott függvény a jövőben is jól fogja leírni az idősor dinamikáját, márpedig egy válság vagy akár egy váratlan pozitív esemény hatására erre nincs semmi biztosíték.
- *Nemparaméteres:* a leggyakoribb egy lineáris szűrő alkalmazása, kevésbé gyakori az exponenciális simítás. Egyre népszerűbb az ún. LOESS simítás, ami egy lokális regressziós függvényillesztést takar.

Lineáris szűrő: $F = \sum_{i=-\infty}^{\infty} a_i B^i$ operátor (a_i -k valós együtthatók), ami a folyamatot "simítja", lényege: a folyamat néhány egymás utáni értékének kiátlagolásával a folyamatban lévő cikcakkok nagyságát jelentősen lecsökkenti, hogy láthatóvá váljon a trend. A lineáris szűrővel gyakran becsülik a trendet: $m_t \approx FX_t = \sum_{i=-\infty}^{\infty} a_i X_{t-i}$. A gyakorlatban rendszerint az itt lévő a_i együtthatók közül csak néhányat választunk 0-tól különbözőnek. A lineáris szűrők speciális esetének tekinthetjük az úgynevezett véges kétoldali **mozgóátlagolást**, amikor $a_i = \frac{1}{1+2q}$, ha $-q \leq i \leq q$ és $a_i = 0$, ha $|i| > q$. Ekkor tehát $\hat{m}_t = \frac{1}{1+2q} \sum_{i=-q}^q X_{t-i}$.

A nemparaméteres megközelítések hátránya, hogy csak lokálisan simítanak, így nem lehet velük közvetlenül előrejelzést készíteni. Amennyiben mégis

szükségünk van előrejelzésekre, akkor meg kell próbálni a simított folyamatra valamilyen jól illeszkedő függvényt illeszteni – azonban nem garantált, hogy sikerül ilyen függvényt találnunk.

2. Trend eliminálása differenciálással – annyiszor alkalmazzuk a ∇ differencia operátort a folyamatra, amíg el nem tűnik a trend. Például lineáris trend esetén már egyszeres differenciálás is eltünteti a trend komponenset.

Definíció. A lineáris szűrő torzítás nélkül enged át egy tetszőleges k fokú polinomot, ha $m_t = \sum_{l=0}^k c_l t^l$ esetén minden t -re $m_t = Fm_t$ teljesül.

Most áttekintjük, hogy mennyivel van több dolgunk, amennyiben a szezonális hatásokkal is kezdeni szeretnénk valamit. A nemstacionárius idősoroknál a trend ÉS a szezonális hatás kimutatására, illetve eltüntetésére az alábbi megközelítéseket lehet követni:

1. Szezonindexek becslése

Legyen x_1, x_2, \dots, x_n a tapasztalati mintánk. Először az előzőekben leírt valamilyen paraméteres vagy nemparaméteres módszerrel kiszűrjük az \hat{m}_t trend hatást, majd az $x_t - \hat{m}_t$ eltérésekből átlagolással megbecsüljük az egyedi szezonhatásokat. Végül ezek átlagával korrigálunk, hogy az összegük 0 legyen. Tehát a két lépés a szezonhatások számszerűsítésére:

- I. Korrigálatlan egyedi szezonindexek becslése (Létrehozunk d darab halmazt, amelyekbe minden d -edik eltérést teszünk be, majd az egyes halmazokban lévő számok átlagát számítjuk. Például az első halmazba tartozik az $1., (d+1)., (2d+1).$ stb. eltérések, a másodikba a $2., (d+2)., (2d+2).$ stb. eltérések):

$$\tilde{s}_k = \frac{\sum_{i: 1 \leq k+id \leq n} (x_{k+id} - \hat{m}_{k+id})}{\sum_{i: 1 \leq k+id \leq n} 1}, \quad k = 1, 2, \dots, d$$

- II. Egyedi szezonindexek korrigálása: $\hat{s}_k = \tilde{s}_k - \frac{1}{d} \sum_{i=1}^d \tilde{s}_i$, $k = 1, 2, \dots, d$

2. Differenciálás – annyiszor alkalmazzuk a ∇ differencia operátort a folyamatra, amíg el nem tűnik a trend. Ezután megnézzük, hogy maradt-e még szezonális a reziduálisokban, és ha igen, akkor alkalmas d -vel alkalmazzuk ∇_d differencia operátort. A d kiválasztásában segítségünkre lehet a folyamat ábrája, illetve az ACF/PACF függvények.
3. Szűrés – alkalmas lineáris szűrővel a trend és a szezonális hatások együttes eliminálása. Polinomiális trend esetén lehet találni ilyen szűrőt.

Most szintetizáljuk az eddigi ismereteinket! Amennyiben az erős időbeli összefüggőséget mutató tapasztalati mintánkat az időtartományban (time domain, szemben a gyakorisági tartománnyal – frequency domain) szeretnénk modellezni, akkor a következő lépéseket kell követni.

Az idősor-modellezés fő lépései (Box-Jenkins modellezés):

1. Az idősor ábrázolása vonaldiagrammal
 - ránézésre homogén, hasonlóan kinéző részekre bontani (amelyek elegendő

mintaelemet tartalmaznak)

- kiugró/hibás/hiányzó értékek kezelése: kihagyás/javítás/békén hagyás
- 2. Előzetes transzformáció, például ha exponenciálisan nő az ábra alapján, akkor érdemes logaritmust venni.
- 3. Trend komponens kiszűrése
- 4. Szezonális komponens kiszűrése
- 5. Modell illesztése – lépései:
 - (i) A megfelelő modellcsalád kiválasztása (a gyakorlatunkon ARMA/ARIMA)
 - (ii) a legjobb modell kiválasztása a modellcsaládon belül valamelyik információs kritérium alapján, például ARMA(1,1)
 - (iii) paraméterbecslés alkalmas módszerrel (momentum, ML-becslés, legkisebb négyzetek)
- 6. Modelldiagnosztika – lépései:
 - (i) becsült együtthatók szignifikánsak-e
 - (ii) a modelltől visszszámolt reziduálisok fehér zaj folyamatot követnek-e
 - (iii) teljesül-e a homoszkedaszticitás. Amennyiben még heteroszkedasztikusak a reziduálisok, akkor illesszünk rájuk GARCH folyamatot.

Amennyiben az illeszkedés megfelelő, akkor mehetünk tovább, ellenkező esetben ugorjunk vissza az 5. pontra vagy a 2. pontra.
- 7. Előrejelzés: általában ez a végső cél, szeretnénk meglévő adataink alapján az idősor jövőbeli viselkedésére minél jobb jóslást adni.

Számos közgazdasági idősor jól modellezhető a következő definícióban szereplő modellcsalád egy alkalmasan választott tagjával.

Definíció. SARIMA (szezonális ARIMA) modell. Az X_t folyamat $SARIMA(p, d, q) \times (P, D, Q)_s$ folyamatot követ s periódussal, amennyiben $Y_t = (1 - B)^d(1 - B^s)^D X_t$ folyamat szezonális $ARMA(p, q)$ folyamatból származik, azaz a következő alakba írható: $\tilde{P}(B)\tilde{R}(B^s)Y_t = Q(B)S(B^s)\varepsilon_t$, ahol

- $\tilde{P}(x) = 1 - \alpha_1 x - \dots - \alpha_p x^p$
- $Q(x) = 1 + \beta_1 x + \dots + \beta_q x^q$
- $\tilde{R}(x) = 1 - \gamma_1 x - \dots - \gamma_r x^r$
- $S(x) = 1 + \delta_1 x + \dots + \delta_Q x^Q$

Idősorok spektrálmélete

Definíció. Legyen X_t stacionárius folyamat, $EX_t = 0$, $\sum_{h=-\infty}^{\infty} |R(h)| < \infty$. Ekkor X_t

spektrális sűrűségfüggvénye: $f(x) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ihx} R(h)$, $-\infty < x < \infty$.

Megjegyzés. Elég $[-\pi; \pi]$ intervallumon megnézni, mert periodikus.

Állítás. A spektrális sűrűségfüggvény tulajdonságai:

- f páros, azaz $f(-x) = f(x) \quad \forall x \in \mathbb{R}$ esetén
- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$ esetén
- $R(h) = \int_{-\pi}^{\pi} e^{ihx} f(x) dx$

Megjegyzés. A spektrális sűrűségfüggvény (1 valószínűséggel) egyértelmű.

Megjegyzés. f Fourier-együtthatói épp az $R(h)$ autokovarianciák.

Állítás. Az $f : [-\pi; \pi] \rightarrow \mathbb{R}$ függvény spektrális sűrűségfüggvénye egy gyengén stacionárius folyamatnak \iff

- f páros;
- $f \geq 0$;
- $\int_{-\pi}^{\pi} f(x) dx < \infty$.

Következmény.

$R(h)$ abszolút szummábilis függvény autokovariancia függvénye egy gyengén stacionárius folyamatnak \iff

- R páros;
- $f(x) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ihx} R(h) \geq 0 \quad \forall x \in [-\pi; \pi]$ esetén.

Definíció. Pozitív szemidefinit függvény. A $g : \mathbb{Z} \rightarrow \mathbb{C}$ függvény pozitív szemidefinit, ha minden $n \in \mathbb{Z}_+$ és minden $t_1, \dots, t_n \in \mathbb{Z}$ esetén a $[g(|t_i - t_j|)]_{i,j=1,\dots,n}$ mátrix pozitív szemidefinit.

Tétel. Az autokovariancia függvény spektrális reprezentációja.

Az $R(h)$ függvény egy gyengén stacionárius folyamat autokovariancia függvénye \iff

$\exists F : [-\pi; \pi] \rightarrow \mathbb{R}$ jobbról folytonos, nemcsökkenő, korlátos, $F(-\pi) = 0$ függvény, amire $R(h) = \int_{-\pi}^{\pi} e^{ihx} dF(x)$.

A fenti tételben lévő F neve: **spektrális eloszlásfüggvény.** Ha

- $\exists f : F(x) = \int_{-\pi}^x f(y) dy$, akkor azt mondjuk, hogy a folyamat *folytonos spektrumú* f spektrális sűrűségfüggvénnyel;
- F tiszta ugrófüggvény, akkor azt mondjuk, hogy a folyamat *diszkrét spektrumú*.

Következmény. $F(\pi) = R(0) = D^2 X_t$, így F nem eloszlásfüggvény (valószínűségi számítási értelemben), mert előfordulhat, hogy $F(\pi) \neq 1$!

A következőkben két általánosabb eredményt mondunk ki.

Tétel. Herglotz-tétel. Legyen $R : \mathbb{Z} \rightarrow \mathbb{C}$ pozitív szemidefinit függvény. Ekkor létezik olyan Q véges mérték $[-\pi; \pi]$ intervallumon, amire $R(h) = \int_{[-\pi; \pi]} e^{ihx} dQ(x)$.

Idősorelméletben a Herglotz-tételt az R -rel jelölt autokovariancia függvényre használjuk, amiről könnyen látható, hogy pozitív szemidefinit. A Herglotz-tételben szereplő Q mértéket **spektrálmértéknek** nevezzük. Ebből a spektrális eloszlásfüggvény: $F(x) = Q([-\pi, x])$, illetve amennyiben Q abszolút folytonos a λ Lebesgue-mértékre nézve, akkor az $f = \frac{dQ}{d\lambda}$ Radon-Nikodym derivált a spektrális sűrűségfüggvény.

Megjegyzés. Ebben a tételben egy mérték szerinti integrál szerepel, míg az egyelő korábbi tételben Riemann-Stieltjes értelemben vett integrál volt.

Egy kis kitekintés: nemcsak a sztochasztikus folyamat autokovariancia függvényének, hanem magának a folyamatnak is van spektrális előállítása, ezt mutatja be a következő tétel.

Tétel. Sztochasztikus folyamat spektrális reprezentációja.

Legyen $T = \mathbb{Z}$, X_k 0 várható értékű gyengén stacionárius folyamat. Ekkor létezik olyan ϕ ortogonális sztochasztikus mérték, amelyre $X_k = \int_{[-\pi; \pi]} e^{ik\mu} d\phi(\mu)$, $k \in \mathbb{Z}$.

Ebben a tételben két dolog is a levegőben lóg:

- mi az a sztochasztikus mérték (a lényege röviden: halmazhoz valószínűségi változót rendel), illetve mikor lesz ortogonális;
 - hogyan definiálható egy ilyen mérték szerinti integrál.
-