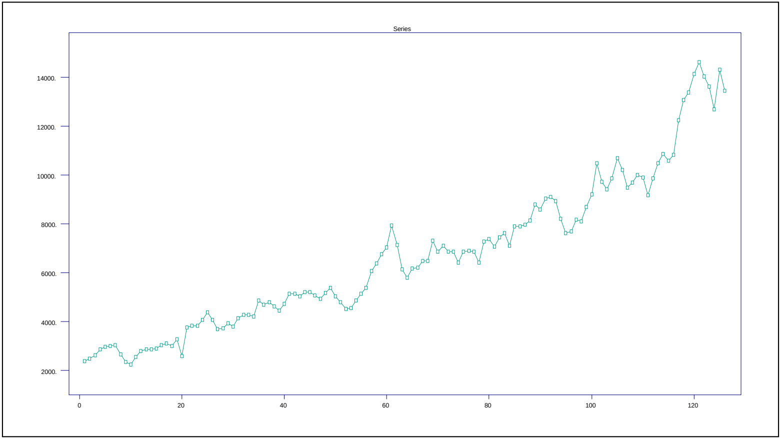
- **Luennoija**: Pentti Saikkonen
- **Tyyppi ja laajuus**: Aineopinnot ja vaativammin suoritettuna syventävät opinnot, 8-10 op
- **Luentoajat**: I ja II periodi, ti 12-14, B120 ja to 12-14, B121
- **Suoritustapa:** Kahdella kurssikokeella tai erilliskokeella yleistentissä; aineopinnot 8 op. Laajennus syventäviin opintoihin 2 op:n harjoitustyöllä.
- **Laskuharjoitukset:** Henri Karttunen
  - I ja II periodi: to 10-12, C122
  - Lisäpisteitä harjoitustehtävien ratkaisemisesta saa 1, 2,...,7, jos ratkaistujen tehtävien määrä on vastaavasti 20, 30, ..., 80 prosenttia.

- **Esitietovaatimukset**
  - Tilastollinen päättely II
  - Lineaariset mallit I
  - Edellisten vaatimat esitiedot
- **Sisältö**
  - Aikasarja-analyysin peruskäsitteet
  - ARMA-mallin ominaisuudet ja sillä ennustaminen
  - ARMA-mallin rakentaminen; mallin valinta, parametrien estimointi ja testaus
  - Finanssiaikasarjojen mallintamisessa käytettävät ARCH- GARCH- ja AR-GARCH-mallit

- **Kirjallisuus ja muu kurssimateriaali**
  - Luentomoniste. Oheislukemistona voi käyttää soveltuvia osia teoksista:
  - P.J. Brockwell & R.A. Davis: Introduction to Time Series and Forecasting. Springer, 1996, 2002 (2nd ed.).
  - J. Hamilton: Time Series Analysis, Princeton University Press, 1994.
  - P.H. Franses & D. van Dijk: Nonlinear Time Series Models in Empirical Finance, Cambridge University Press, 2000 (finanssiaikasarjoista).
  - H. Lütkepohl & M. Krätzig (toim.): Applied Time Series Econometrics, Cambridge University Press, 2004. (JMulti-ohjelmistoon liittyvä kirja)
- **Muuta**
  - Opiskeltavia malleja voi soveltaa esimerkiksi R:llä (koodit saatavissa kurssisivulta) tai ilmaiseksi saatavissa olevalla JMulTi-ohjelmistolla (http://www.jmulti.de/).

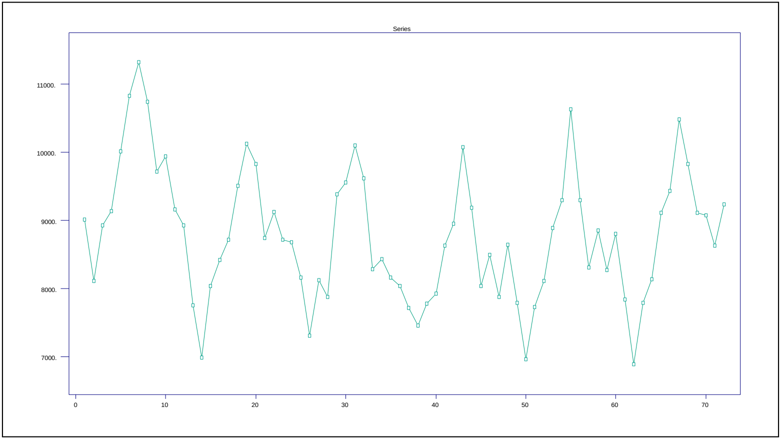
Johdanto

- Aikasarja on havaintoaineisto, jossa havainnot on saatu peräkkäisinä ajankohtina: y_t, $t = 1, ..., T$, jossa t on ajankohta ja T havaintojen lukumäärä.
- Esimerkiksi vuosiaineistossa $t = 1970, ..., 2005$ tai kuukausiaineistossa $t =$ 1970I, 1970II,...,2005XII.
- Tällä kurssilla oletetaan, että peräkkäiset ajankohdat ovat ajallisesti yhtä kaukana toisistaan tai niitä voidaan käsitellä sellaisina.
- Jos y_t on skalaari, puhutaan *yksiulotteisesta tai reaaliarvoisesta aikasarjasta.* Jos y_t on vektori, puhutaan *moniulotteisesta tai vektoriarvoisesta aikasarjasta*.
- Aikasarja-aineistoja analysoidaan useilla sovellusalueilla kuten taloustieteissä ja muissa yhteiskuntatieteissä, luonnontieteissä, insinööritieteissä ja lääketieteessä.



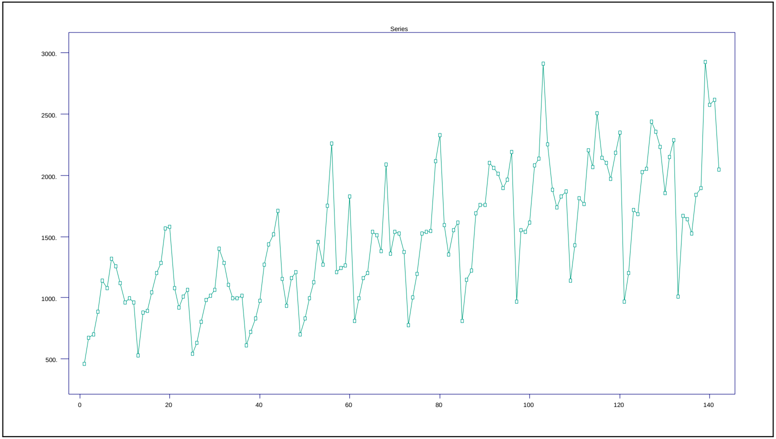
Kuvio 1.1. Australian tavaroiden ja palvelusten neljännesvuosittainen tuonti ajanjaksolta 1959III - 1990IV (miljoonaa Australian dollaria).

- Ominaista nouseva *trendi*.



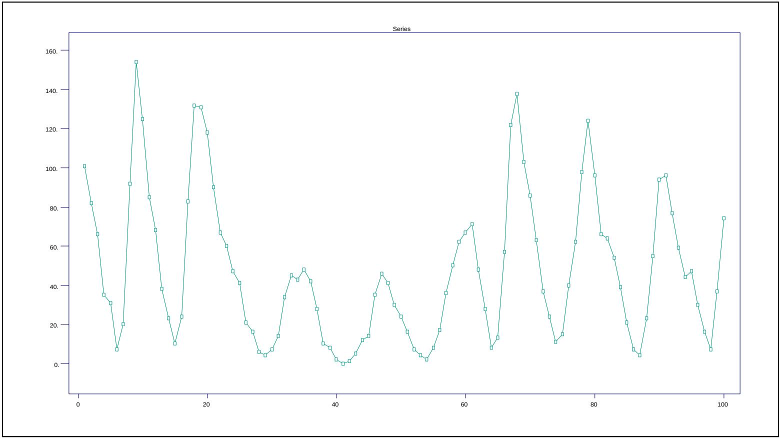
Kuvio 1.2. USA:ssa auto-onnettomuuksissa kuukausittain kuolleiden lukumäärä ajanjaksolta 1973I - 1978XII.

- Ei selvää nousevaa tai laskevaa trendiä.
- Sen sijaan selvä *kausivaihtelu*. Vuotuinen maksimi aina heinäkuussa ja minimi helmikuussa.



Kuvio 1.3. Australialaisen punaviinin kuukausittaisen myynnin volyymi (tuhansina litroina) ajanjaksolta 1980I - 1991X.

- Sekä nouseva trendi että selvää kausivaihtelua.
- Vuotuinen huippu yleensä heinäkuussa, myynti alhaisimmillaan tammikussa. Myynnin vaihtelu kasvaa myynnin määrän kasvaessa (tai ajan kuluessa).



Kuvio 1.4. Vuotuisten auringonpilkkujen lukumäärä vuosilta 1770 - 1870.

- Ei trendiä eikä kausivaihtelua.
- *Syklistä* vaihtelua, jossa syklin luonne vaihtelee eri aikoina: Peräkkäisten huippujen välinen etäisyys ei ole vakio.
- Peräkkäisten havaintojen välillä positiivinen korrelaatio: Suuret havaintoarvot esiintyvät yleensä peräkkäin ja samoin pienet.

Johdanto

- **Aikasarja-analyysin tavoitteena** on usein rakentaa *tilastollinen malli*, jolla kuvataan havaittua aikasarjaa (aikasarjoja) ja sen (niiden) vaihtelua sekä peräkkäisten havaintojen välistä, yleensä ilmeistä, riippuvuutta.
- **Kun malli on valittu, voidaan**
  - estimoida sen parametrit
  - tarkistaa estimoidun mallin ja havaintojen yhteensopivuus
  - testata parametreja koskevia hypoteeseja
  - käyttää mallia ajateltuun tarkoitukseen.

Johdanto

- **Mallia voidaan käyttää esimerkiksi**
  - antamaan havaintoaineistosta *tiivistetty kuvaus*, joka auttaa ymmärtämään aineiston taustalla olevaa ilmiötä
  - *ennustamaan* aikasarjan (aikasarjojen) tulevia arvoja - moniulotteisessa tapauksessa kysymyksessä on yhden aikasarjan ennustaminen muiden (selittävien) aikasarjojen avulla
  - tutkimaan moniulotteisessa tapauksessa aikasarjojen välisten ajallisten *riippuvuuksia* - esim. punaviinin myynnin ja valkoviinin myynnin välistä riippuvuutta

Johdanto

- Ensiksi on aina syytä piirtää tarkasteltava aikasarja.
- Tällöin saadaan käsitys aikasarjan pääpiirteistä kuten trendistä ja/tai kausivaihtelusta ja/tai ilmeisistä äkillisistä muutoksista tai poikkeavista havainnoista, joita ei seuraavassa oleteta olevan.
- Monissa aikasarjamalleissa trendiä ja/tai säännöllistä kausivaihtelua ei oleteta olevan.
- Jos niitä on ja tällaisia malleja halutaan soveltaa, käytetään muunnettua aikasarjaa, jossa valitulla muunnoksella eliminoidaan nämä aikasarjan piirteet.
- Alkuperäisen aikasarjan kannalta katsoen trendi ja/tai kausivaihtelu mallinnetaan tällöin muunnoksella, joka voi vaatia parametrien estimointia.

- Lineaarisen trendin tapauksessa voidaan havaittuun aikasarjaan sovittaa pienimmän neliösumman (PNS) menetelmällä trendisuora ja siirtyä tarkastelemaan residuaalisarjaa.
- Havaittu aikasarja y_t korvataan tällöin aikasarjalla

$$\hat{x}_t = y_t - \hat{\alpha} - \hat{\beta} t, \quad t = 1, ..., T,$$

jossa estimaatit $\hat{\alpha}$ ja $\hat{\beta}$ on laskettu PNS-menetelmällä. Lineaarisen funktion asemesta voidaan yleisempiä polynomeja.
- Kausivaihtelun mallintamiseen voidaan käyttää trigonometrisia ajan funktioita tai ns. *kausi-indikaattoreita*, jolloin esim. neljännesvuosiaineistossa käytetään residuaaliaikasarjaa

$$\hat{x}_t = y_t - \hat{\beta}_1 d_{1t} - \cdots - \hat{\beta}_4 d_{4t}, \quad t = 1, ..., T,$$

jossa i. kausi-indikaattori d_{it} saa arvon yksi, kun kysymyksessä on neljänneksen i havainto ja nolla muulloin ($i = 1, ..., 4$).

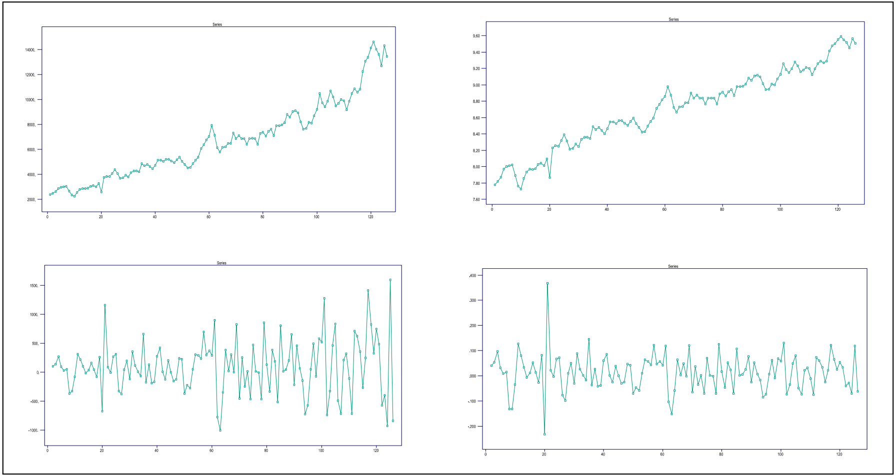
- Muutokset eli differenssit:

$$y_t - y_{t-s}, \quad s \geq 1.$$

  Tapaus $s = 1$ tavallisin; kausivaihtelun yhteydessä voi myös olla $s =$ kausivaihtelujakson pituus (esim. $s = 4$ tai $s = 12$).
- Usein differensointiin yhdistetään logaritmointi (olettaen $y_t > 0 \ \forall t$).
- Logaritmointia suositaan erityisesti, kun aikasarjassa on eksponentiaalinen trendi tai kun aikasarjan vaihtelu kasvaa tason kasvaessa.
- Logaritmisten differenssit voidaan tulkita suhteellisten muutosten indikaattoreiksi, sillä

$$\log y_t - \log y_{t-1} \approx \left(y_t - y_{t-1}\right) / y_{t-1},$$

  kun muutos on pieni.



Kuvio 1.5. Australian tavaroiden ja palvelusten tuonti.

- Alkup. sarja (vas. ylh.), saman logaritmi, jossa trendi lähes lineaarinen (oik. ylh.) ja näiden differenssit (alh.), jossa vasemmalla vaihtelu kasvaa sarjan tason kasvaessa.

Johdanto

- Kurssilla tarkastellaan malleja, jotka olettavat, että havaitussa aikasarjassa ei ole trendiä tai että trendi on eliminoitu muunnoksella ja käytetään muunnettua sarjaa.
- Kurssilla ei tarkastella moniulotteisia aikasarjamalleja eikä malleja, joissa trendin mallinnus olisi varsinaisen mielenkiinnon kohteena.

- **Lähtökohta**: Tarkoituksena on rakentaa tilastollinen malli havaitulle yksiulotteiselle aikasarjalle $y_1, ..., y_T$, jossa ei ole trendiä.
- Hyvin yleisellä tasolla havaittu aikasarja tulkitaan tällöin satunnaisvektorin $y = (y_1, ..., y_T)$ havaituksi arvoksi eli *realisaatioksi*.
- **Huom.:** Samaa merkintää käytetään satunnaisille suureille ja niiden havaituille arvoille. Ero ymmärrettävä asiayhteydestä.

Stationaariset prosessit

Peruskäsitteitä

- Mallinnuksessa spesifioidaan havaitun aikasarjan "tuottaneen" satunnaisvektorin $y = (y_1, ..., y_T)$ todennäköisyysjakauma.
- Havainnot eivät riippumattomia $\Rightarrow$ komponenttien y_t $(t = 1, ..., T)$ reunajakaumien spesifiointi ei riitä.
- Konkreettisissa tilanteissa y:n yhteisjakauman spesifiointi tapahtuu yleensä spesifioimalla peräkkäisten havaintojen riippuvuutta kuvaava malliyhtälö, johon liitetään "sopiva" jakaumaoletus.

- Matemaattisissa tarkasteluissa ja myös aikasarjamalleja määriteltäessä on kätevää tarkastella havaittua aikasarja vastaavan satunnaisvektorin $y = (y_1, ..., y_T)$ "laajennusta".

- Diskreettiaikainen *stokastinen prosessi* (usein vain prosessi) on joukko satunnaismuuttujia

$$\{y_t; t = 0, \pm 1, \pm 2, ...\}$$

tai toisinaan

$$\{y_t; t = 0, 1, 2, ...\} \quad \text{tai} \quad \{y_t; t = 1, 2, ...\}\,.$$

- Koska aikaindeksin arvojoukko on yleensä tiedossa, voidaan merkitä lyhyesti y_t tai $\{y_t\}$, jos halutaan korostaa eroa prosessin yksittäiseen komponenttiin.

- Käytännön kannalta on merkityksellistä, että prosessista $\{y_t; t = 0, \pm 1, \pm 2, ...\}$ on käytettävissä vain yksi havaittu realisaatio eli yksi havainto $(y_1, ..., y_T)$.
- Lisäksi, prosessista $\{y_t; t = 0, \pm 1, \pm 2, ...\}$ saatu havainto havaitaan vain osittain eli ajankohtina $t = 1, ..., T$.
- Edellä sanotusta seuraa, ettei tarkasteltavan prosessin ominaisuuksia voida selvittää käytettävissä olevan havaintoaineiston $y_1, ..., y_T$ avulla ilman prosessia $\{y_t; t = 0, \pm 1, \pm 2, ...\}$ koskevia rajoittavia oletuksia.

- Tarkastellaan prosessin y_t *odotusarvofunktiota*

$$\mu_t = \mathrm{E}\left(y_t\right), \quad t = 0, \pm 1, \pm 2, \ldots$$

ja *kovarianssifunktiota*

$$\gamma_{s,t} = \mathrm{Cov}\left(y_s, y_t\right) = \mathrm{E}\left[\left(y_s - \mu_s\right)\left(y_t - \mu_t\right)\right], \quad t = 0, \pm 1, \pm 2, \ldots\ .$$

- Näiden funktioiden estimointi yhden realisaation avulla mahdotonta, ellei niiden riippuvuutta ajasta rajoiteta jotenkin.
- Oletaan, ettei tällaista ajallista riippuvuutta (tietyssä mielessä) ole.

- Prosessi y_t on *heikosti stationaarinen tai kovarianssistationaarinen*, jos

$$\mathsf{E}\left(y_t\right) = \mu \quad \text{kaikilla} \quad t = 0, \pm 1, \pm 2, \ldots$$

ja

$$\mathsf{Cov}\left(y_t, y_{t+h}\right) = \gamma_{t,t+h} = \gamma_{0,h} \quad \text{kaikilla } h, \; t = 0, \pm 1, \pm 2, \ldots \; .$$

- Ts., pätee *aikainvarianttius*: Odotusarvofunktio μ_t on ajassa vakio ja kovarianssifunktio $\gamma_{s,t}$ ei riipu ajankohdista s ja t, vaan pelkästään niiden välisestä etäisyydestä $t - s$. Merkitään lyhyesti $\gamma_{0,h} = \gamma_h$. Erityisesti,

$$\gamma_0 = \mathsf{Var}\left(y_t\right).$$

- γ_h:ta kutsutaan y_t:n *autokovarianssikertoimeksi* viipymällä h ja h:n funktiona y_t *autokovarianssifunktioksi*. μ:tä nimitetään *odotusarvoksi*.

- Autokovarianssifunktiolla on ominaisuudet

$$\gamma_0 \geq 0, \quad |\gamma_h| \leq \gamma_0 \quad \text{ja} \quad \gamma_h = \gamma_{-h}.$$

  Tapaus $\gamma_0 = 0$ mielenkiinnoton $\Rightarrow$ oletetaan $\gamma_0 > 0$.

- Käytännössä käytetään yleensä *autokorrelaatiofunktiota*

$$\rho_h = \mathrm{Cor}\,(y_t, y_{t+h}) = \gamma_h / \gamma_0.$$

- *Autokorrelaatiokertoimilla* ρ_h on ilmeiset ominaisuudet

$$\rho_0 = 1, \quad |\rho_h| \leq 1 \quad \text{ja} \quad \rho_h = \rho_{-h}.$$

- Sekä autokovarianssifunktiota että autokorrelaatiofunktiota riittää siis tarkastella vain viipymilla $h \geq 0$ (jälkimmäistä viipymillä $h > 0$).

- Heikon stationaarisuuden takaama aikainvarianttius $\Rightarrow$ $\mathrm{E}\left(y_t\right) = \mu$ ja $\mathrm{Cor}\left(y_t, y_{t+h}\right) = \gamma_h$ ilmeisesti mahdollista estimoida.
- Ongelmana vielä, että autokorrelaatiofunktiossa on yleisesti ottaen ääretön määrä estimoitavia suureita, joita kaikkia ei voida käytännössä estimoida.
- Yleensä voidaan kuitenkin *olettaa*, että

$$\gamma_h \to 0, \quad \text{kun} \quad h \to \infty.$$

- Tällöin satunnaismuuttujat y_t ja y_{t+h} ovat lähes korreloimattomia, kun h on "suuri" eli kun y_t ja y_{t+h} ovat ajallisesti "kaukana" toisistaan.
- Käytännössä riittää siten estimoida autokorrelaatiofunktio ρ_h, kun $h = 1, ..., H$ ja H on niin suuri, että $\rho_h \approx 0$, kun $h > H$.

- Oletetaan, että y_t on *heikosti stationaarinen.*

  Tällöin $\mathrm{E}\left(y_1\right) = \cdots = \mathrm{E}\left(y_T\right) = \mu$

  ja luonteva odotusarvon estimaattori on otoskeskiarvo

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t .$$

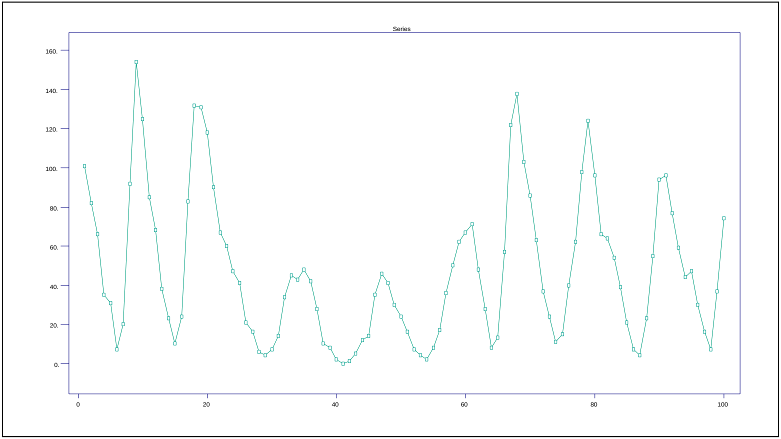
- Tällöin on myös $\mathrm{Cov}\left(y_1, y_{1+h}\right) = \cdots = \mathrm{Cov}\left(y_{T-h}, y_T\right) = \gamma_h$ $(h \geq 0)$ ja $\mathrm{E}\left(y_t\right) = \mathrm{E}\left(y_{t+h}\right)$,

  joten luonteva autokovarianssikertoimen γ_h estimaattori on

$$\mathrm{c}_h = \frac{1}{T-h} \sum_{t=1}^{T-h} \left(y_t - \bar{y}\right)\left(y_{t+h} - \bar{y}\right), \quad 0 \leq h < T,$$

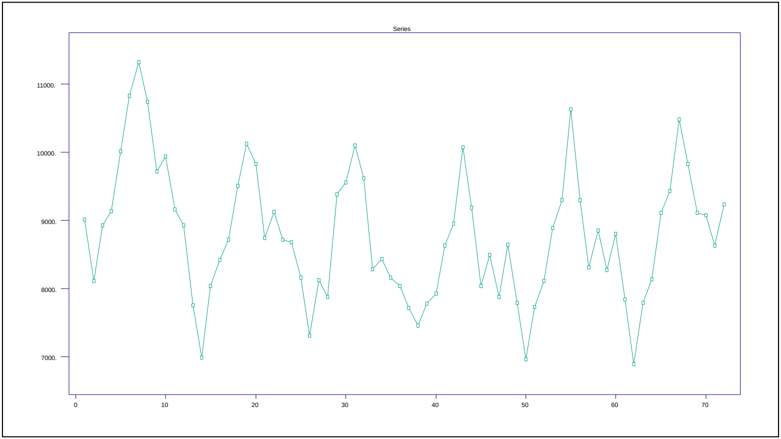
  ja luonteva autokorrelaatiofunktion estimaattori on siten

$$\mathrm{r}_h = \mathrm{c}_h / \mathrm{c}_0, \quad 0 \leq h < T.$$



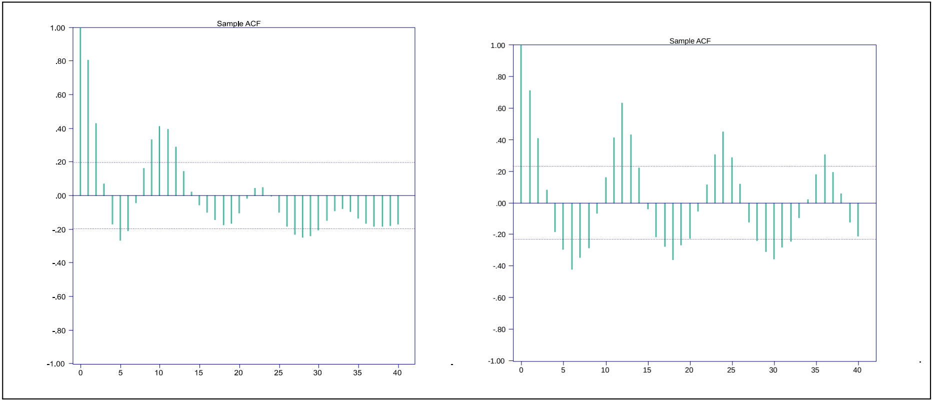
Kuvio 1.4. Vuotuisten auringonpilkkujen lukumäärä vuosilta 1770 - 1870.

- Ei trendiä eikä kausivaihtelua.
- *Syklistä* vaihtelua, jossa syklin luonne vaihtelee eri aikoina: Peräkkäisten huippujen välinen etäisyys ei ole vakio.
- Peräkkäisten havaintojen välillä positiivinen korrelaatio: Suuret havaintoarvot esiintyvät yleensä peräkkäin ja samoin pienet.



Kuvio 1.2. USA:ssa auto-onnettomuuksissa kuukausittain kuolleiden lukumäärä ajanjaksolta 1973I - 1978XII.

- Ei selvää nousevaa tai laskevaa trendiä.
- Sen sijaan selvä *kausivaihtelu*. Vuotuinen maksimi aina heinäkuussa ja minimi helmikuussa.



Kuvio 2.1. Auringonpilkkusarjan otosautokorrelaatiofunktio (vasemmalla) ja onnettomuussarjan otosautokorrelaatiofunktio (oikealla) viipymillä $h = 0, ..., 40$.

- Yksittäinen estimaattori r_h kuuluu likimain 95%:n todennäköisyydellä kuvioihin piirrettyjen rajojen väliin, jos $\rho_h = 0$ kaikilla $h > 0$

- **Esimerkki 2.1.**(i) Prosessi ε_t on *(heikko) valkoinen kohina*, jos

$$\mathrm{E}\left(\varepsilon_t\right) = \mu, \quad \mathrm{Var}\left(\varepsilon_t\right) = \sigma^2 < \infty \quad \text{ja} \quad \mathrm{Cov}\left(\varepsilon_s, \varepsilon_t\right) = 0, \quad \text{kun} \quad s \neq t.$$

- $\mathrm{E}\left(\varepsilon_t\right)$ on t:stä riippumaton vakio ja

$$\mathrm{Cov}\left(\varepsilon_t, \varepsilon_{t+h}\right) = \begin{cases} \sigma^2, & \text{kun} \quad h = 0 \\ 0, & \text{kun} \quad h \neq 0, \end{cases}$$

joten heikko stationaarisuus on määritelmän nojalla ilmeinen.

- Yksinkertaisuudestaan huolimatta keskeinen prosessi, sillä sen avulla voidaan määritellä monimutkaisempia prosesseja.

- **Esimerkki 2.1.**(ii) Määritellään prosessi

$$y_t = A\cos(\lambda t) + B\sin(\lambda t), \quad \lambda \in [0, \pi) \text{ vakio,}$$

jossa satunnaismuuttujat A ja B totetuttavat

$$\mathrm{E}(A) = \mathrm{E}(B) = 0, \quad \mathrm{Var}(A) = \mathrm{Var}(B) = \sigma^2 \text{ ja } \mathrm{Cov}(A, B) = 0.$$

- Tällöin

$$\mathrm{E}(y_t) = 0 \text{ ja } \mathrm{Cov}(y_t, y_{t+h}) = \sigma^2 \cos(\lambda h) \text{ riippuu vain } h\text{:sta,}$$

joten y_t on heikosti stationaarinen (perustelu jätetään tehtäväksi).

- Realisaatiot $\cos(\lambda t)$:n ja $\sin(\lambda t)$:n lineaarikombinaatioita.
- Huomaa, että ehto $\gamma_h \to 0$, kun $h \to \infty$ ei päde.

- Heikko stationaarisuus on sikäli havainnollinen käsite, että sen realistisuutta voidaan arvioida katsomalla havaitun aikasarjan kuvaa.
- Jos havainnot vaihtelevat vakiovarianssilla kiinteän tason ympärillä, voidaan heikkoa stationaarisuutta pitää kohtuullisena oletuksena.
- Heikko stationaarisuus ei kuitenkaan riitä aikasarjamallien tilastolliseen analysointiin, sillä se ei yksistään mahdollista uskottavuusfunktion muodostamista eikä estimaattorien ja testisuureiden jakaumien tutkimista.
- Tästä syystä määritellään heikkoa stationaarisuutta vahvempi stationaarisuuden muoto.

- Prosessin $\{y_t; t = 0, \pm 1, \pm 2, ...\}$ sanotaan olevan *vahvasti stationaarinen*, jos satunnaismuuttujilla

$$y_{t_1}, ..., y_{t_m} \qquad \text{ja} \qquad y_{t_1+h}, ..., y_{t_m+h}$$

  on sama $m-$ulotteinen yhteisjakauma kaikilla kokonaisluvuilla $t_1, ..., t_m$, h ja m $(m > 0)$.

- Ts., prosessin prosessin y_t koko todennäköisyysstruktuuri (ei vain ensimmäiset ja toiset momentit) on aikainvariantti.

- Jos y_t on vahvasti stationaarinen, niin

$$\left(y_{t_1}, \ldots, y_{t_m}\right) \sim \left(y_{t_1+h}, \ldots, y_{t_m+h}\right) \text{ kaikilla } t_1, \ldots, t_m, h \text{ ja } m \ (m > 0)\,.$$

- Pätee siis,

  VS1 Satunnaismuuttujat y_t ovat samoin jakautuneita $\forall\ t$.

  VS2 Satunnaisvektorit $\left(y_t, y_{t+h}\right)$ ja $\left(y_s, y_{s+h}\right)$ ovat samoin jakautuneita kaikilla t ja s ja jokaisella (kiinteällä) h.

  VS3 y_t on heikosti stationaarinen, jos $\mathrm{E}\left(y_t^2\right) < \infty$.

Stationaariset prosessit

Stokastisen prosessin vahva stationaarisuus

- Prosessi y_t (ei välttämättä missään mielessä stationaarinen) on *normaalinen*, jos satunnaismuuttujien $y_{t_1}, \ldots, y_{t_m}$ $m-$ulotteinen todennäköisyysjakauma on normaalinen kaikilla kokonaisluvuilla $t_1, \ldots, t_m$ ja m $(m > 0)$.
- Koska ensimmäiset ja toiset momentit määräävät normaalijakauman täysin, seuraa heikosta stationaarisuudesta ja prosessin normaalisuudesta vahva stationaarisuus.
- Yleisesti heikosta stationaarisuudesta ei seuraa vahva stationaarisuus (heikko stationaarisuus ei esim. rajoita toista kertalukua korkeampia momentteja).
- Vahvasta stationaarisuudesta ei seuraa heikko stationaarisuus, jos $\mathrm{E}\left(y_t^2\right) = \infty$ (vrt. esim. Cauchyn jakauma).

Stationaariset prosessit

Stokastisen prosessin vahva stationaarisuus

- Vahva stationaaruus (toisin kuin heikko) säilyy funktiomuunnoksissa:

  VS4 Jos z_t on vahvasti stationaarinen prosessi, niin samoin on prosessi

  $$y_t = g\left(z_{t+m}, \ldots, z_{t-n}\right), \quad m, n \geq 0,$$

  millä tahansa ”siistillä” (eli esim. jatkuvalla tai yleisemmin ns. mitallisella) funktiolla g. Lisäksi m:n tai n:n tai molempien paikalla voi olla ∞.

- Kun m ja n ovat äärellisiä seuraa ominaisuus VS4 jokseenkin suoraan vahvan stationaarisuuden määritelmästä.
- Ellei stationaarisuuden luonnetta täsmennetä, tulkitaan stationaarisuus jatkossa vahvaksi stationaarisuudeksi (ei yleisesti käytössä oleva tulkinta).

- **Esimerkki 2.2.**(i) Olkoon ε_t jono riippumattomia ja samoin jakautuneita satunnaismuuttujia kertymäfunktiona $F(\cdot)$.
- Tämä prosessi on vahvasti stationaarinen ja sitä kutsutaan *vahvaksi valkoiseksi kohinaksi*.
- Jos $\mathrm{E}\left(\varepsilon_t^2\right) < \infty$, on ε_t myös heikko valkoinen kohina eli pätee

$$\mathrm{E}(\varepsilon_t) = \mu, \quad \mathrm{Var}(\varepsilon_t) = \sigma^2 < \infty \quad \text{ja} \quad \mathrm{Cov}(\varepsilon_s, \varepsilon_t) = 0, \quad \text{kun} \quad s \neq t.$$

- Kun $\mathrm{E}\left(\varepsilon_t^2\right) < \infty$, käytetään tästä prosessista lyhennystä

$$\varepsilon_t \sim \mathrm{iid}\left(\mu, \sigma^2\right),$$

jossa $\mu = \mathrm{E}(\varepsilon_t)$ ja $\sigma^2 = \mathrm{Var}(\varepsilon_t)$ (jatkossa yleensä $\mu = 0$).

- Jos edellisessä oletetaan lisäksi $\varepsilon_t \sim \mathrm{N}\left(\mu, \sigma^2\right)$, merkitään

$$\varepsilon_t \sim \mathrm{nid}\left(\mu, \sigma^2\right).$$

- **Esimerkki 2.2.**(ii) Määritellään prosessi

$$y_t = \varepsilon_t \sqrt{\omega + \alpha \varepsilon_{t-1}^2}, \quad \omega > 0, \quad \alpha \geq 0 \quad \text{ja} \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right).$$

- Laskemalla nähdään, että y_t on heikko valkoinen kohina.
- Koska $y_{t-1} = \varepsilon_{t-1} \sqrt{\omega + \alpha \varepsilon_{t-2}^2}$, on selvää, ettei y_t ole tapausta $\alpha = 0$ lukuun ottamatta vahva valkoinen kohina.
- Heikosta valkoisesta kohinasta, jolle siis pätee

$$\mathsf{E}\left(\varepsilon_t\right) = \mu, \quad \mathsf{Var}\left(\varepsilon_t\right) = \sigma^2 < \infty \quad \text{ja} \quad \mathsf{Cov}\left(\varepsilon_s, \varepsilon_t\right) = 0, \quad \text{kun} \quad s \neq t,$$

käytetään merkintää $\varepsilon_t \sim \text{wn}\left(\mu, \sigma^2\right)$ (wn $\leftrightarrow$ 'white noise').

- Valkoisen kohinan avulla määritellä monimutkaisempia prosesseja.
- Yksinkertainen esimerkki on *ensimmäisen asteen liukuvan keskiarvon prosessi* eli MA(1)-prosessi (MA $\leftrightarrow$ 'moving average')

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right), \tag{2.2}$$

- Prosessin saamien arvojen oletetaan syntyvän kahden riippumattoman ja ei-havaittavan satunnaissokin painotettuna summana.
- MA(1)-prosessi on vahvasti stationaarinen (ks. Esimerkki 2.2(i) ja ominaisuus VS4) ja myös heikko stationaarisuus pätee.
- Ilmeinen yleistys on MA(q)-prosessi

- Laskemalla nähdään, että MA(1)-prosessin

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right), \tag{2.2}$$

1. ja 2. momentit ovat (myös oletuksella $\varepsilon_t \sim \text{wn}\left(0, \sigma^2\right)$)

$$\mathrm{E}\left(y_t\right) = 0, \quad \mathrm{Var}\left(y_t\right) = \sigma^2\left(1+\theta^2\right)$$

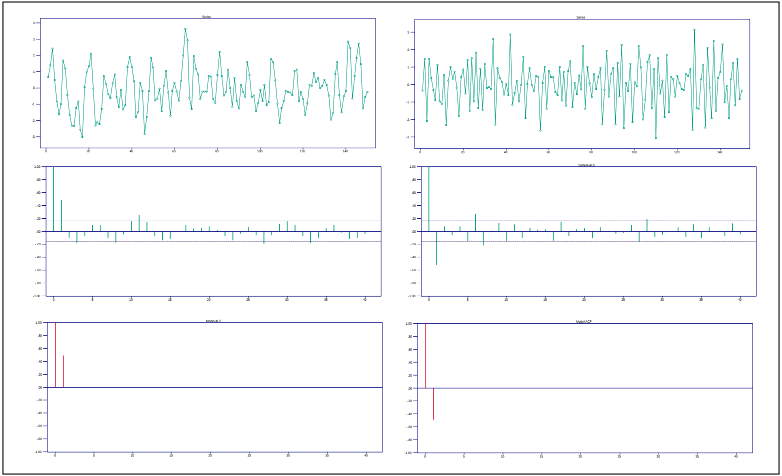
ja

$$\mathrm{Cov}\left(y_t, y_{t+h}\right) = \begin{cases} \theta\sigma^2, & h = 1 \\ 0, & h > 1 \text{ (tai } h < -1\text{).} \end{cases}$$

- Autokorrelaatiofunktio on siten

$$\rho_h = \begin{cases} 1, & h = 0 \\ \theta/\left(1+\theta^2\right), & h = 1 \\ 0, & h > 0. \end{cases}$$

- MA(1)-prosessille on siis ominaista, että autokorrelaatiofunktiossa on katkos viipymällä yksi.



Kuvio 2.2. Kaksi MA(1)-prosessista $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\varepsilon_t \sim \text{nid}\,(0,1)$, simuloitua aikasarjaa ($T = 150$), niiden otosautokorrelaatiofunktiot (keskellä) ja teoreettiset autokorrekaatiofunktiot (alinna) viipymillä $h = 0, ..., 40$. Vasemmalla $\theta = 0.8$ ja oikealla $\theta = -0.8$.