Meiri spektral-teoría ofl.

Helgi Tómasson
helgito@hi.is

30. janúar 2024

- Gerum ráð fyrir að stationary ferli með meðaltal 0 sé mælt á 4 tímapunktum. Gögnin eru mælingar á $\boldsymbol{X}' = (X_1, X_2, X_3.X_4)$. Tek (discrete)-Fourier transform í tíðnunum, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ þar sem $\lambda_j = 2\pi j/4$ þar sem $j = 1, \ldots, 4$, þ.e. $\lambda_1 = 2\pi/4 = \pi/2$, $\lambda_2 = 4\pi/4 = \pi$, $\lambda_3 = 6\pi/4 = 3\pi/2$, $\lambda_4 = 8\pi/4 = 2\pi$.
- Fourier-transformið er:

$$g(\lambda_j) = \frac{1}{4}\sum_{t=1}^{4} \exp(-i\lambda_j t)X_t$$

- Á fylkjaformi:

$$\underbrace{\frac{1}{4}\begin{bmatrix} -i & -1 & i & 1 \\ -1 & 1 & -1 & 1 \\ i & -1 & -i & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{A} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} g(\lambda_1) = \tilde{X}_1 \\ g(\lambda_2) = \tilde{X}_2 \\ g(\lambda_3) = \tilde{X}_3 \\ g(\lambda_4) = \tilde{X}_4 \end{bmatrix}$$

- Það því búið að skrifa mælivektorinn $\boldsymbol{X}$ í öðru hnitakerfi:

$$A\boldsymbol{X} = \tilde{\boldsymbol{X}}$$

- Breytileikinn, varíansinn í $\boldsymbol{X}$ og $\tilde{\boldsymbol{X}}$ er sá sami. Einstök hnit í $\tilde{\boldsymbol{X}}$ segja til um það hvaða tíðnir eru mikilvægar.
- Það má nálga ferlið með því að velja ákveðnar mikilvægar tíðnir og nota síðan andhverju A til að varpa til baka.
- Þetta er svipað og að nálga fall með margliðum.
- Rétt skalað $V(\tilde{\boldsymbol{X}})$ er spektur fyrir ferlið.
- Fallið symmsetrískt og því bara notaður helmingur af tíðnunum.
- Mjög reiknifrekt og því nota nútímamenn *fft*, fast-fourier-transform.
- Ferlið $\boldsymbol{X}$ einkennist af sjálffylgni, ferlið $\tilde{\boldsymbol{X}}$ einkennist af misdreifni.
- Í stórum úrtökum eru fylgni milli hnita $\tilde{X}$. Athugið að ganga verður út frá því að sjálffylgni $\boldsymbol{X}$ sé „cirular".

Sýnidæmi um ARMA líkön

- Upprifjun um diffurjöfnur. Ef

$$y'' + ay' + by = 0, \qquad \text{þá er ákveðin margliða skoðuð}$$

$$\text{ef rætur eru komplexar þá er lausnarferill sveiflukenndur}$$

- Sama á við um ARMA líkön:

$$\phi(L)Y_t = \theta(L)\varepsilon_t$$

- Ef rætur $\phi(z)$ eru komplexar þá verður ferlið y_t sveiflukennt.
- Ef spektral þéttifallið:

$$f(\lambda) = \frac{\sigma^2}{2\pi}\frac{|\theta(exp(-i\lambda)|^2}{|\phi(exp(-i\lambda)|^2}$$

hefur topp í punktinum λ_0 þá er mikill breytileiki vegna sveiflu af lengdinni $2\pi/\lambda_0$.

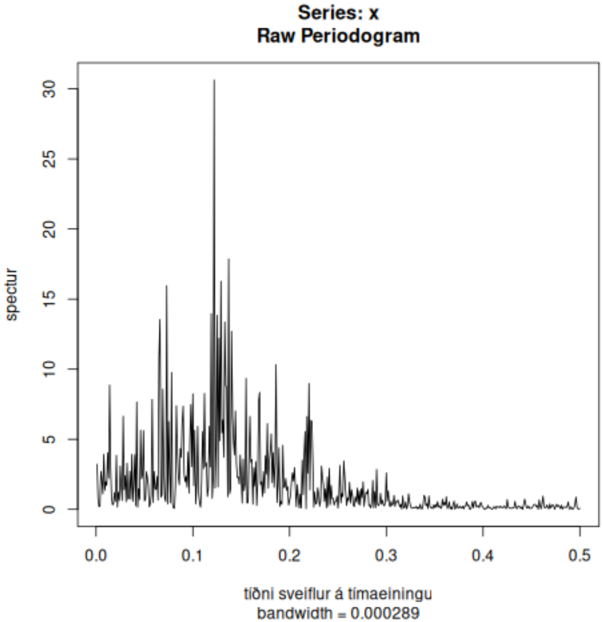
- AR(2) er einfaldasta líkan sem getur skýrt sveiflu. T.d.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

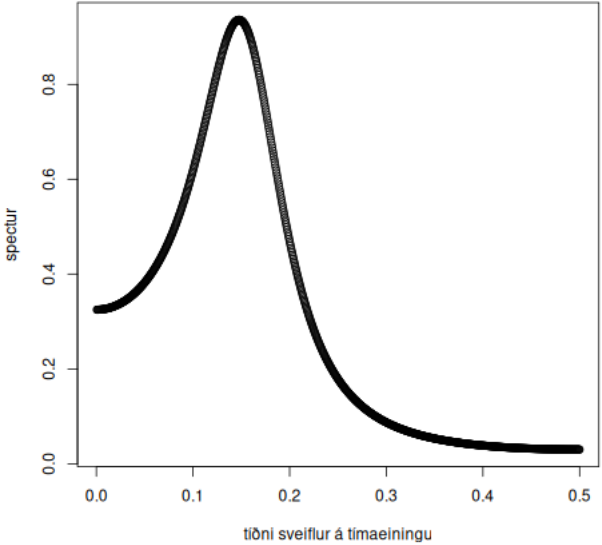
```
phi1=0.8
phi2=-0.5
sigma=1
ii=sqrt(as.complex(-1))
y=arima.sim(n=1000,list(ar=c(phi1,phi2),sd=sigma))

lam=0
specfunc=function(lam){
ff=sigma^2/(2*pi)*1/abs((1-phi1*exp(ii*lam)-
phi2*exp(ii*lam*2)))^2
return(ff)
}
```

- Plotta myndir:

Úrtaksspectur, $\hat{f}(\lambda)$



Fræðilegt spectur, ϕ_1 og ϕ_2 þekkt



Hugleiðingar

- Úrtaksspektrið, $\hat{f}(\lambda)$ hefur topp í tíðninni 0.122 sem samsvarar sveiflu upp á 1/0.122=8.2 tímaeininga sveiflu.
- Fræðilega spectrið $f(\lambda)$ hefur topp í tíðninni 0.146 sem samsvarar sveiflu upp á 1/0.146=6.78 tímaeininga sveiflu. Berið saman við formúlu í bók, ca. bls 45.
- Væri hægt að nota $\hat{f}(\lambda)$ til að giska á ϕ_1, ϕ_2 og σ?
- Já. Það er hægt að framkvæma einhvers konar regression:

$$\hat{f}(\lambda_j) = f(\lambda_j) + u_j, \quad \text{þar sem } V(u_j) \simeq f(\lambda_j)^2.$$

- Margar útfærslur til af þessari hugmynd. Þetta er stundum kallað Whittle-likelihood aðferð, eftir Whittle frá ca. 1950. Einnig hugmynd er lýst í bókinni.

Um long-memory, kafli 2.11

- Höfum hingað til rætt um stutt minni, þ.e. fortíðin afskrifast með veldishraða. Spektralfall fyrir slík ferð er endanleg tala í tíðninni 0.
- Í stationary ferlum eru bönd á hve stór framtíðar óvissa getur verið.
- Í random-walk vex framtíðar óvissa stöðugt (í hlutfalli við hversu langt skal spá).
- Er til eitthvað þarna á milli?
- Já. Við höfum ferli þar sem sjálffylgnifallið deyr hægar út en exponential.
- Auðskiljanlegast er ARFIMA, þ.e. fractionally-integrated ARIMA, ARIMA(p,d,q) þar sem er brot.

ARFIMA

- Beran (1994) er lykilheimild varðandi langtímaminni. Þar segir að ferli sýni langtíma minni ef spektralfallið er óendanlegt í að minnsta kosti einni tíðni. Það jafngildir því að sjálffylgnifallið sé ekki tölugildissummanlegt.
- Meðfærilegur flokkur líkana sem nær þeim eiginleika eru ARFIMA(p,d,q),
  $(1-L)^d\Phi(L)Y_t = \Theta(L)\varepsilon_t$, þar sem d er leyft að vera brot.
- Hugmyndin er að vinna í and Box-Jenkins (1970) að láta ϕ_i og θ_i ná skammtímasveiflum og nota d sem mælikvarða á minnið.
- Sjálffylgnifall ARFIMA deyr út polynomialt, þ.e. mun hægar en með veldishraða. Áhrif fortíðar deyja mjög hægt út.
- Stærðfræðilega eru svona ferli ekki Markov né martingale.
- Spectralfall ARFIMA(p,d,q) er á forminu:

$$f(\lambda) = \frac{\sigma^2}{2\pi}\underbrace{|(1-\exp(i\lambda)|^{-2d}}_{long-memory}\frac{|\Theta(i\exp(\lambda)|^2}{|\Phi(i\exp(\lambda))|^2}.$$

Ýmsir eiginleikar

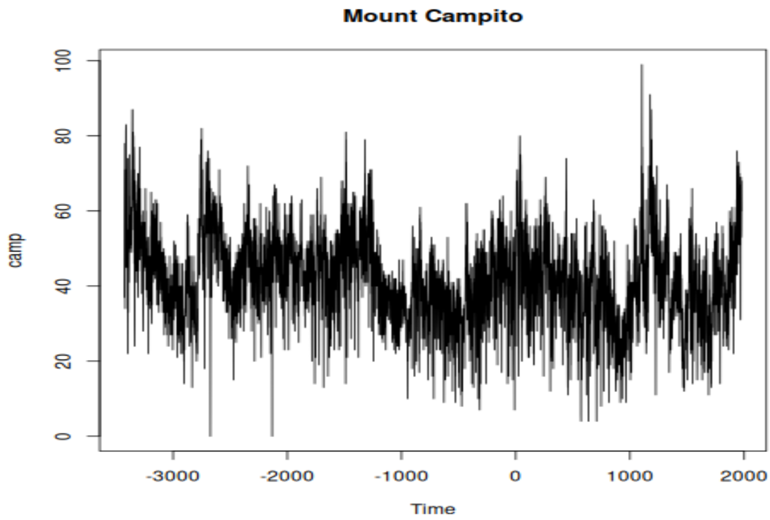
- ARFIMA(p,d,q) með $0 < d < 1/2$ hefur þann eiginleika að $f(0) = \infty$, $\int f(\lambda)d\lambda = VAR(Y(t))$ og auto-covariancar $\gamma(k)$.
- Vegna þess hve mælingar eru háðar verða ályktanir um t.d. meðaltal $\mu = E(Y(t))$ og þýðingu ytri stærða frábrugðnar hefðbundinni ályktanafræði.
- Flóknar formúlur gilda um dreifingu úrtaksmeðaltals. Staðalfrávik þess er stundum ekki $\propto T^{-1/2}$ heldur $\propto T^{-1/2+d}$. Þ.e. það er erfitt að álykta um meðaltal.
- Með því að sigta 0-tíðni frá og nota spektral aðferðir getur verið auðveldara að álykta um d en μ.

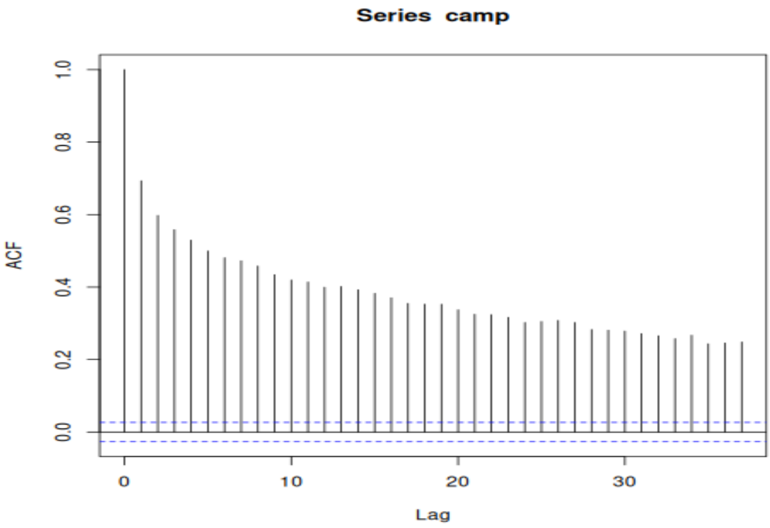
Ýmsar útvíkkanir og hagnýtingar

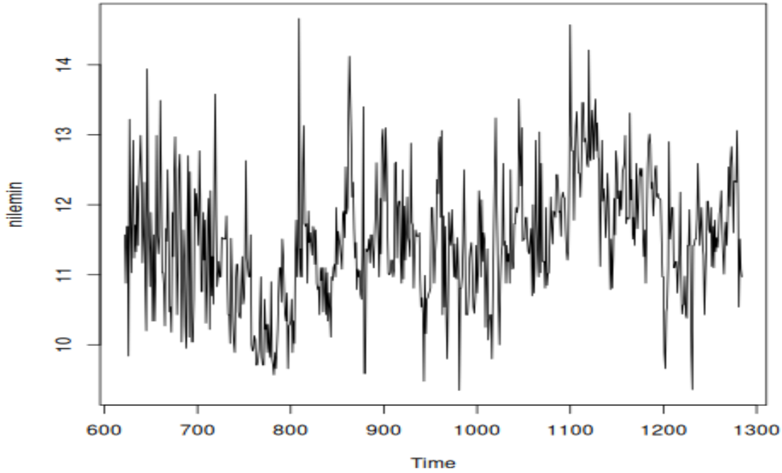
- Baillie ofl. (2002) skoða árhringi úr trjám.
- Chan ofl.(1998) sýna state-space nálgun, sbr. kafla 11 í bók
- Comte og Renault (1996) sýna hvernig megi skilgreina fractional Wiener-ferli (sjá kafla 6) í samfelldum tíma í mörgum víddum.
- Dagsvik ofl. (2020) meta long-memory raðir fyrir mælt hitastiga á ca. 100 veðurstöðvum í 200 ár ásamt öðrum hitastigsgögnum í ca. 2000 ár.
- Ýmsar nálganir (því líkönin eru stærðfræðilega og reiknifræðilega erfið), t.d. FerARMA, með eiginleika sem minna á long-memory, $\Phi(L)^{\delta_1} Y_t = \Theta(L)^{\delta_2} \varepsilon_t$.

Uppruninn

- Hurst (1951)tók saman nokkur hundruð ára gögn um flóðahæð í Níl og sá að sjálffylgnifallið dó hægt út. Hann fann upp hugtakið Hurst exponent=d+1/2 (fyrir normaldreifingu??).
- Granger ofl. (1980) rekja hvernig long-memory ferli getur myndast sem summa margra AR(1) ferla. Hugmynd sem ýmsir hafa nýtt sér. T.d. að líkanið sé summa margrar AR(1) ferla.


Mount Campito
camp
0
20
40
60
80
100
-3000
-2000
-1000
0
1000
2000
Time


Series camp
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
Lag



Mynd: Lágmarksvatnshæð í Níl

Almennt um long-memory

- Hvernig á að giska á d?
  1. Spektral aðferðir þægilegar, Whittle aðferð.
  2. Exact aðferðir, maximum-likelihood erfiðar vegna reiknitíma. Durbin-levinson algtoritmi virkar ef mælt er reglulega í tíma.
  3. Vegna erfiðleika, hentugt að vera með nálganir, ARTFIMA, FerARMA (fractional-unit-root), ofl.
  4. Þó fræðin séu erfið eru margir pakkar í t.d. R sem hægt er að nota, arfima, artfima, o.s.frv.
  5. Þeir sem skilja vandann hafa því mörg tæknitól aðgengileg.
  6. Long-memory kemur fyrir í fjármálum og náttúruvísindum.