OBLIGATORISK OPPGAVE 1

STA-2003-Tidsrekker

16. februar 2019

Martin Soria Røvang
Universitetet i Tromsø

Inneholder 12 sider, inkludert forside.

Innhold

1

1.1 a

Hvit støy er en prosess som består av ukorrelerte tilfeldige variabler med forventing $\mu_w = 0$ og varians σ_w^2.

$$W_t \sim WN(0, \sigma_w^2) \tag{1}$$

En hvit støy prosess man ofte kommer over er den Gaussiske hvite støy prosessen gitt ved,

$$W_t \sim N(0, \sigma_w^2) \tag{2}$$

1.2 b

Her har vi at $\hat{\rho}(h) \overset{d}{\sim} \mathcal{N}(0, 1/n)$, og dermed standardavvik $\sigma_{\hat{\rho}} = 1/\sqrt{n}$. Grensen der 95% av verdiene er innenfor vil da være 1.96 standardavvik, slik at $a = -1.96/\sqrt{n}$ og $b = 1.96/\sqrt{n}$.

$$\mathscr{P}(-1.96/\sqrt{n} < \hat{\rho}(h) < 1.96/\sqrt{n}) = 95\%. \tag{3}$$

Dette fordi,

$$P(Z > 1.96) = 0.025$$

der Z er standardnormalfordeling $Z = \frac{x-\mu}{\sigma}$, her er $\mu = 0$, siden ACF har fordelingen $\mathcal{N}(0, 1/n)$ setter man inn får vi

$$P(\frac{x}{\sigma} > 1.96) = 0.025$$

løser for x gir,

$$P(x > 1.96\sigma) = 0.025$$

der sannsynligheten for x over 1.96σ er 2.5%. Normalfordeling er symmetrisk rundt forventingen og vi vil derfor ha det samme for,

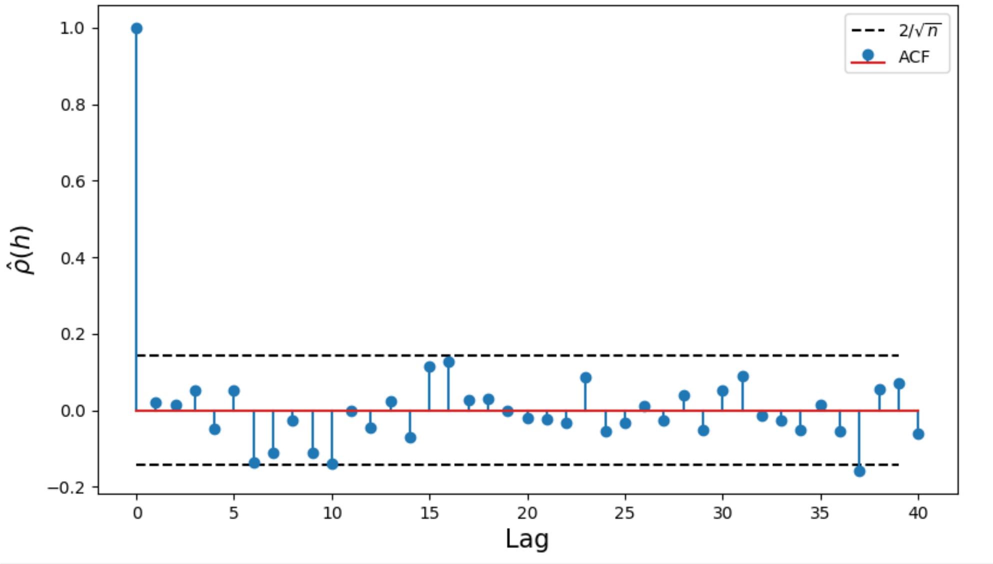
$$P(x < -1.96\sigma) = 0.025$$

Derfor vil verdier mellom dette intervallet gi $1 - (0.025 + 0.025) = 0.95$

Ved bruk av python får vi samme resultat,

```
from scipy.stats import norm
ss = norm.cdf(1.96) - norm.cdf(-1.96)
print(ss)
>> 0.95
```

I figur(1) under kan man se verdiene på ACF over lag,



Figur 1: ACF til en hvit støy prosess

Her kan man se at nesten alle verdiene ligger mellom ± 1.96 standardavvik.

1.3 c

Tidsrekke 1 er *ikke konsistent* med hvit støy prosess fordi tidlige verdier er sterkt korrelert med de fremtidige.
Tidsrekke 2 er *konsistent* med hvit støy prosess fordi korrelasjonen med tidligere verdier er veldig svak og tilfeldig.
Tidsrekke 3 er ikke hvit støy prosess fordi lag 1 er korrelert med tidligere verdier.

2

2.1 a

En stokastisk prosess er Gaussisk hvis hvert endelig datapunkt for hver endelig tid er en tilfeldig variabel X som følger en multidimensjonal gaussisk distribusjon. Derfor vil en Gaussisk prosess ha stokastisk variabel $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Gamma})$,

$$f_{\boldsymbol{X}}(x_1,..,x_N;t_1,..,t_N) = \frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Gamma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{X}})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{X}})\right], \tag{4}$$

hvor prosessen $\boldsymbol{x} = [x_1,..,x_N]^T$, gjennomsnittsvektoren$\bar{\boldsymbol{X}} = [E[X_1],..,E[X_N]]^T$ og kovariansmatrisen er $\boldsymbol{\Gamma} = E[(\boldsymbol{x}-\bar{\boldsymbol{X}})(\boldsymbol{x}-\bar{\boldsymbol{X}})^T]$ der$|\cdot|$ angir determinanten.

2.2 b

Vet at $E[x_t] = 0$, hvis vi ser på varianser får vi,

$$Cov(x(t), x(s=t)) = \frac{\sigma^2}{2}(t^{2H} + t^{2H} - |t-t|^{2H}) = \sigma^2 t^{2H}$$

Her ser vi at variansen er avhengig av tid og prosessen er dermed ikke stasjonær. Hvis variansen ikke hadde vært avhengig av tid måtte vi fortsatt se om autokovariansen kun var avhengig av *lag* $|s-t| = h$ da variansen kun er et spesialtilfelle der $s = t$.

2.3 c

Har at $Cov(y_t, y_s) = E[y_t y_s] - E[y_t]E[y_s]$, der $E[y_t] = E[y_s] = 0$. Vi har da at,

$$Cov(X(t), Xs) = E[X(t)X(s)] = \frac{\sigma^2}{2}(t^{2H} + s^{2H} - |t-s|^{2H})$$

legger vi inn for $y(t)$ og $y(s)$ får vi $E[(x(t)-x(t-1))(x(s)-x(s-1))]$,

$$E[(x(t)-x(t-1))(x(s)-x(s-1))] = E[x(t)x(s) - E[x(t-1)x(s)] - E[x(s-1)x(t)] + E[x(t-1)x(s-1)]$$

løser opp og får,

$$E[y_t y_s] = \frac{\sigma^2}{2}(t^{2H} + s^{2H} - |t-s|^{2H} - ((t-1)^{2H}$$
$$+s^2 - |(t-1)-s|^{2H})) - (t^{2H} + (s-1)^{2H} - |t-(s-1)|^{2H}))$$
$$+((t-1)^{2H}(s-1)^{2H} - |(t-1)-(s-1)|^{2H})$$

rydder vi opp får vi,

$$E[y_t y_s] = \frac{\sigma^2}{2}(-|t-s|^{2H} + |t-1-s|^{2H} - |t-s+1|^{2H} - |t-1-s+1|^{2H})$$

trekker sammen ledd,

$$Cov(y_t, y_s) = E[y_t y_s] = \frac{\sigma^2}{2}(|t-s+1|^{2H} + |t-1-s|^{2H} - 2|t-s|^{2H})$$

Variansen blir da,

$$Cov(y_t, y_{s=t}) = E[y_t y_t] = \frac{\sigma^2}{2}(|t-t+1|^{2H} + |t-1-t|^{2H} - 2|t-t|^{2H})$$

som blir,

$$\gamma(0) = \frac{\sigma^2}{2}(1+1-0) = \sigma^2$$

forenkler,

$$Cov(y_{t+h}, y_t) = \gamma(h) = \frac{\sigma^2}{2}(|h+1|^{2H} + |h-1|^{2H} - 2|h|^{2H})$$

og tilslutt auto-korrelasjonsfunksjonen(ACF) $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$,

$$\rho(h) = \frac{1}{2}(|h+1|^{2H} + |h-1|^{2H} - 2|h|^{2H})$$

som er det vi ville vise.

2.4 d

Vet at, $H = 1/2$, $E[y_t] = 0$ da har vi,

$$\gamma(h) = \frac{\sigma^2}{2}(|h+1| + |h-1| - 2|h|) \tag{5}$$

dette gir auto-kovariansfunksjonen gitt ved *lag* h,

$$\gamma(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

og auto-korrelasjonsfunksjonen(ACF) $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$,

$$\rho(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

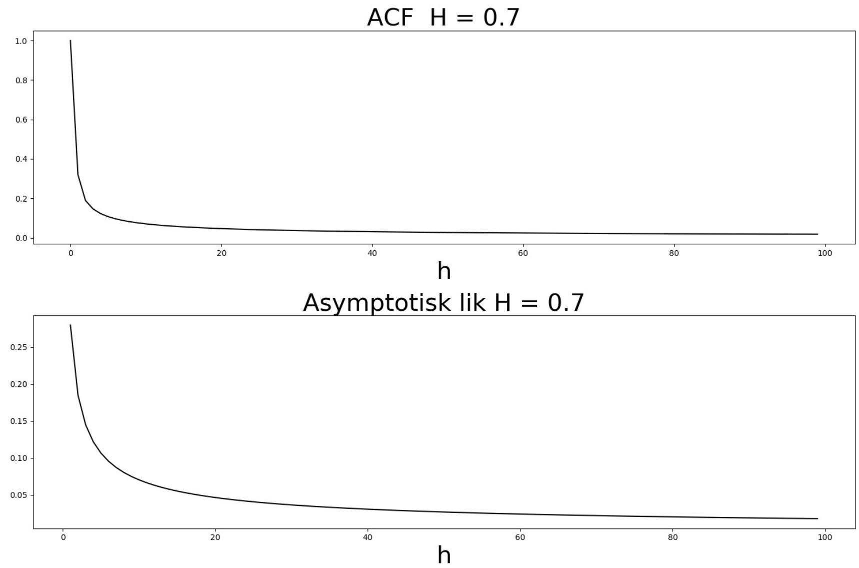
Denne følger en hvit støy prosess.

2.5 e

Denne prosessen er strengt stasjonær da vi vet at dette er en gaussisk prosess(iid). Hvis den ikke hadde vært iid (identically independent distributed) hadde den vært svakt stasjonær fordi prosessen er avhengig av *lag* h.

2.6 f

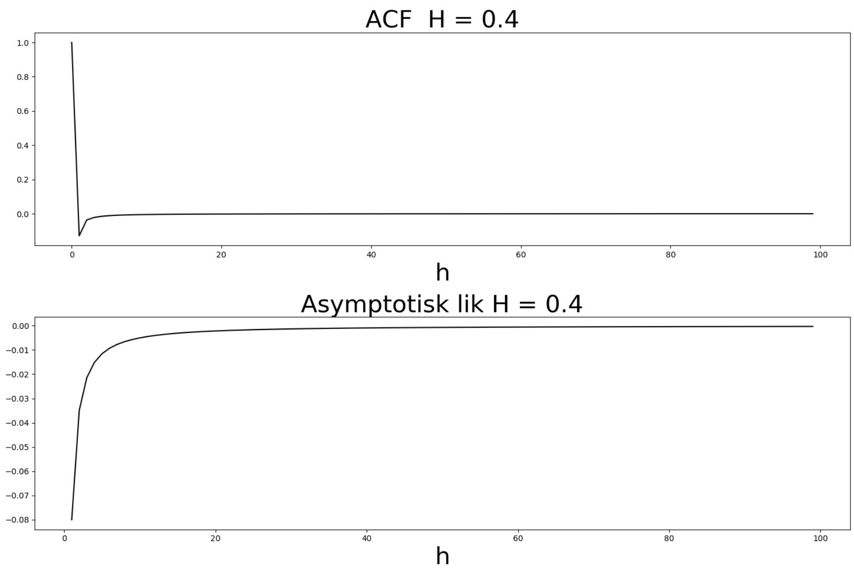
I figur(2) under er ACF plottet med den asymptotiske funksjonen.



Figur 2: ACF og den som er asymptotisk like funksjonen

Her ser man at når $h->\infty$ blir de mer lik. Her er $H=0.7$ og her er det positiv men minkende korrelasjon.

I figur(3) er det samme plottet med verdi $H=0.4$



Figur 3: ACF og den som er asymptotisk like funksjonen, $H=0.4$

Her ser man at når $h->\infty$ blir de mer lik som tidligere, men med negativ korrelasjon i starten untatt $h=0$.

Ved sammenligningstesten så konvergerer summen $\sum_0^\infty \rho(h)$, der $H>1/2$.

2.7 g

Har,

$$Cov(X(at),X(at))=\frac{\sigma^2}{2}((at)^{2H}+(at)^{2H}-(at-at)^{2H}) \tag{6}$$

løser opp og får,

$$Cov(X(at),X(at))=\sigma^2(at)^{2H} \tag{7}$$

for $a^HX(t)$ har vi,

$$Cov(a^HX(t),a^HX(t))=a^{2H}Cov(X(t),X(t))=\sigma^2(at)^{2H} \tag{8}$$

Vi har da $[X(at)...X(at_n)]\sim[a^HX(t)...a^HX(t_n)]$

3

3.1 a

Starter med den lineære prosessen med støy som har 0 forventning og varians σ_w^2

$$x_t = \beta_0 + \beta_1 t + w_t \tag{9}$$

Forventing av x_t blir,

$$E[x_t] = E[\beta_0 + \beta_1 t + w_t] = \beta_0 + \beta_1 t$$

fordi $\beta_0 + \beta_1 t$ leddene ikke er stokastiske og at forventning av det stokastiske leddet w_t er null.

Siden forventning er avhengig av tid er prosessen ikke stasjonær.

3.2 b

Vi har differansen i første orden $\nabla^1 x_t$ gitt ved,

$$\nabla x_t = x_t - x_{t-1} = (\beta_0 + \beta_1 t + w_t) - (\beta_0 + \beta_1 (t-1) + w_{t-1})$$

forkorter vi,

$$\nabla x_t = \beta_1 + w_t - w_{t-1}$$

Forventning blir da,

$$E[\nabla x_t] = E[\beta_1 + w_t - w_{t-1}] = \beta_1$$

Den er da ikke avhengig av tid, men vi må også se om autokovariansen er avhengig av tid, her kan vi se om prosessen er svakt stasjonær $\approx$ stasjonær (kun avhengig av tidsdifferansen/lag h).

$$\gamma_x(h) = E[(\nabla x_{t+h} - \mu)(\nabla x_t - \mu)]$$

her er $\mu = \beta_1$

$$\gamma_x(h) = E[(\nabla x_{t+h} - \beta_1)(\nabla x_t - \beta_1)] \tag{10}$$

Legger inn og får,

$$\gamma_x(h) = E[(\beta_1 + w_{t+h} - w_{t+h-1} - \beta_1)(\beta_1 + w_t - w_{t-1} - \beta_1)] \tag{11}$$

Forenkler,

$$\gamma_{\nabla x_t} = E[(w_{t+h} - w_{t+h-1})(w_t - w_{t-1})] \tag{12}$$

dette blir,

$$\gamma_{\nabla x_t} = E[w_{t+h}w_t - w_{t+h-1}w_t - w_{t+h}w_{t-1} + w_{t+h-1}w_{t-1}] \tag{13}$$

dette gir oss en stykkevis funksjon,

$$\gamma_{\nabla x_t} = \begin{cases} 2\sigma_w^2, & h = 0 \\ -\sigma_w^2, & |h| = 1 \\ 0, & \text{eller.} \end{cases}$$

Her er auto-kovariansfunksjonen kun avhengig av tidsdifferansen h, derfor er denne prosessen stasjonær.

3.3 c

Bytter vi ut støy leddet med en stasjonær prosess y_t har vi forventningen,

$$E[\nabla x_t] = E[\beta_1 + y_t - y_{t-1}] = \beta_1$$

fordi $E[y_t] = E[y_{t-1}] = \mu_y$.
Autokovariansen blir,

$$\gamma_{\nabla x_t} = E[(x_{t+h} - x_{t+h-1} - \mu)(x_t - x_{t-1} - \mu)]$$

igjen er $\mu = \beta_1$. forenkler man dette får man,

$$\gamma_{\nabla x_t} = E[(y_{t+h} - y_{t+h-1})(y_t - y_{t-1})]$$

som blir,

$$\gamma_{\nabla x_t} = E[y_{t+h}y_t - y_{t+h-1}y_t - y_{t+h}y_{t-1} + y_{t+h-1}y_{t-1}]$$

forenkler,

$$\gamma_{\nabla x_t} = E[y_{t+h}y_t] - E[y_{t+h-1}y_t] - E[y_{t+h}y_{t-1}] + E[y_{t+h-1}y_{t-1}]$$

$$\gamma_{\nabla x_t} = \gamma(h) - \gamma(h-1) - \gamma(h+1) + \gamma(h) = 2\gamma(h) - \gamma(h-1) - \gamma(h+1)$$

Denne prosessen er derfor stasjonær fordi autokovariansen er avhengig av *lag* h.

4 Appendix

5 Referanser