. előadás 2010. XI. 30.

Csüt 17^{15} – 18^{00} is konzultáció

Statisztika

1., Becsléselmélet

2., Hipotézis~~elmélet~~vizsgálat → valami igaz-e, vagy nem.

1, Becsléselmélet

- pontbecslések → pl. a fénycső exp ideig ég → erre becslünk
- intervallumbecslés → konfidenciaintervallum, amibe nagy eséllyel esik (nagy valószínűséggel ide esnek az adatok.

Pontbecslések

~~Időegységenként~~

Van egy fergeteges szerver, 10 időegységben megmérték a beérkezések (sessionok) számát.

234, 123, 240, 235, 235, 197, 178, 260, 251, 197 :

ezt találták.

Tudjuk, hogy független Poisson-eloszlásúak ϑ paraméterrel.

$$\underline{x} = (x_1, x_2, \dots, x_{10}) =$$

Ez a $\underline{x}$ egy megvalósulása az $X_1, X_2, \dots, X_{10}$ független mintának, ahol ezek az X-ek valószínűségi változók.

Az $X_1, X_2, \dots, X_{10}$ Poisson-eloszlásúak ϑ paraméterrel, és ez a ϑ nem ismert.

A minta alapján határozzuk meg ϑ-t!

Általában: (a feladat):

Adott P_ϑ $\vartheta \in \Theta$ eloszlások egy halmaza

Θ: paramétertér

Az előbbi példában $P_\vartheta = Poi(\vartheta)$

↓

$$P_\vartheta(X=k) = \frac{\vartheta^k}{k!} e^{-\vartheta}$$

~~$\Theta \ni (0, 1, \infty)$~~ $\Theta = (0, \infty)$

↖ tetszőleges 0 és ∞ közötti számot felvehet.

Θ-ból valahogy kiválasztódik (nem mi választjuk!) egy paraméter: ϑ. Ez az igazi paraméter.

A P_ϑ eloszlásból generáltunk egy (független) mintát:

$x_1, x_2, \dots, x_n$.

A megvalósulás: $\underline{x} = (x_1, x_2, \dots, x_n)$

Feladat: Találjuk meg ϑ-t!

Def: (statisztika): Statisztikának nevezzük a mintaelemek egy tetszőleges függvényét.

↓

Példák:

$\overline{X}_n$: átlag $= \dfrac{x_1 + \dots + x_n}{n}$ $\left(\dfrac{\text{mintaelemek összege}}{\text{mintaelemek száma}}\right)$

A példában $\overline{x}_{10} = 2{,}5$

tapasztalati szórásnégyzet: $S_n^2 = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2$

A példában $S_{10}^2 = 1732$.

Ez eddig mérhető $\mathbb{R}^n \to \mathbb{R}$ függvény.

Rendezett minta: $\left(x_1^*, x_2^*, \dots x_n^*\right)$, ahol

x_i^* a minta i-edik legnagyobb eleme.

A példában: ez $= (260, 251, \dots, 123)$

Ez $\mathbb{R}^n \to \mathbb{R}^n$ függvény.

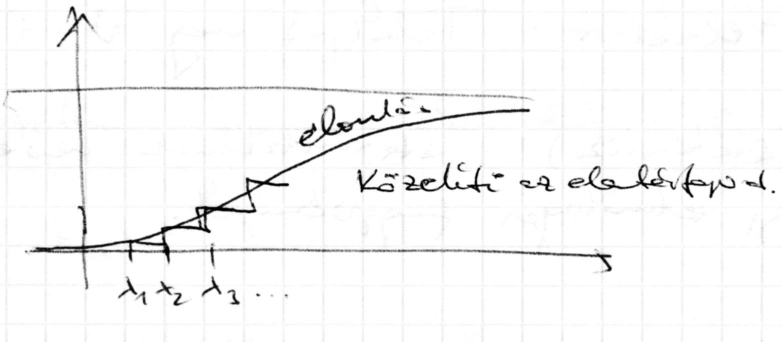
~~Ez az eloszlásfüggvény~~

Az eloszlásfgv becslésére a tapasztalati eloszlásfüggvény:

$$F_n^*(x) = \frac{1}{n} \# \left\{ i : X_i < x \right\}$$

ott ugrik, ahol van egy mintaelem, s ott $\frac{1}{n}$-et ugrik.

Ha pl:



$\mathbb{R}^n \to$ valós fgv-ek tere

Ez nem olyan jó becslés. (Inkább paramétert becslünk).

Becslés; mitől jó egy becslés?

Adott egy eloszláscsalád: P_ϑ, $\vartheta \in \Theta$.

Adott egy $\psi(\vartheta)$ fgv, ahol ψ rögzített.

Ezt szeretnénk becsülni. ($\psi(\vartheta)$-t)

↓

szeretnénk megmondani az igazi ϑ mellett.

$m(\vartheta)$ a P_ϑ eloszlású valószínűségi változó várható értéke.

A példában: $m(\vartheta) = \vartheta$

ahol $P_\vartheta = Poi(\vartheta)$

Ha viszont $P_\vartheta = EXP(\vartheta)$, akkor $m(\vartheta) = \frac{1}{\vartheta}$

↑ itt a várható értéket akarjuk becsülni.

<u>Def(1)</u>: Egy $T(\underline{X})$ statisztika torzítatlan becslése a $\psi(\vartheta)$-nak, ha a várható értéke:

$$E_\vartheta\left(T(\underline{X})\right) = \psi(\vartheta) \qquad \forall \vartheta \in \Theta$$

<u>Áll</u>: (Konkrét ψ-re):

A mintaátlag $T(\underline{X}) = \overline{X}_n = \frac{X_1 + \dots + X_n}{n}$ torzítatlan becslése a várható értéknek. ($m(\vartheta)$).

↑

veszünk n elemű mintát (n sok). pl 1000db 10 elemű. Mindegyikből vesszük ezt a statisztikát (átlagot), s ezek átlaga a várható érték.

A mindegyik statisztikák átlaga az az érték lesz, amit becsülni szeretnénk.

<u>Biz</u>: $E_\vartheta\left(T(\underline{X})\right) =$ $\quad X_1, X_2, \dots, X_n \sim P_\vartheta$ ← ebből generálódott

$$= E_\vartheta\left(\frac{X_1+\dots+X_n}{n}\right) = \frac{1}{n}\left(\underbrace{E_\vartheta X_1 + \dots + E_\vartheta X_n}\right) = E_\vartheta X_1 = m(\vartheta)$$

ami E_ϑ miatt van, mert P_ϑ eloszlású

Az $m(\vartheta)$ pedig egy P_ϑ eloszlású valószínűségi változó várható értéke.

(A becslésünk nagyon rossz nem lehet).

Ellenpélda: S_n^2

$\uparrow$

a tapasztalati szórásnégyzetet

a szórást kellene becsülni

$$E_\vartheta(S_n^2) = \frac{n-1}{n}\underbrace{\sigma^2(\vartheta)}$$

P_ϑ eloszlású valószínűségi változó szórásnégyzete.

Ez egy torzított becslése a szórásnégyzetnek.

A példában:

$\frac{X_1+\dots+X_{10}}{10} = 21{,}5$ egy torzítatlan becslése az igazi paraméternek.

Def: Aszimptotikusan torzítatlan, ~~ha $n\to\infty$~~ ha $X_1, X_2, \dots$ minta esetén $E_\vartheta(T(\underline{X}_n))$ $\quad \underline{X}_n = (X_1, X_2, \dots, X_n)$

Def: T aszimptotikusan torzítatlan becslése $\psi(\vartheta)$-nak, ha $X_1, X_2, \dots$ minta esetén $\lim_{n\to\infty} E_\vartheta(T(\underline{X}_n)) = \psi(\vartheta)$

Pl: A tapasztalati szórásnégyzet S_n^2 aszimptotikusan torzítatlan becslése $\sigma^2(n)$-nek: $\frac{n-1}{n} \to 1$, ha $n \to \infty$.

Torzítatlanság: fix n-re a becslés a keresett paraméter körül van.

Def(2): (Konzisztencia)

A $T(X_n)$ statisztika-sorozat az $X_1, X_2, \ldots$ mintán konzisztens becslése $\Psi(\vartheta)$-nak, ha

$$~~\lim~~ \; T(\underline{X}_n) \xrightarrow[n\to\infty]{} \Psi(\vartheta)$$

Ahogy növeljük a mintaelemszámot, válaszolódik az igazi paraméter valamilyen fgv-ére.

Például: Poi eloszlásnál: $P_\vartheta = \mathrm{Poi}(\vartheta)$

és az $\overline{X}_n$ mintaátlag konzisztens becslése az $m(\vartheta)$-nak.

Mivel $\mathrm{Poi}(\vartheta) \Rightarrow m(\vartheta) = \vartheta$, ezért a ϑ igazi paraméternek is konzisztens becslése.

Ez igaz, mert: $\frac{X_1 + X_2 + \ldots + X_n}{n} \to E_\vartheta(X_1) = m(\vartheta)$

az átlag konvergál a várható értékhez a nagy számok törvénye alapján.

Tehát Poisson-eloszlás esetén a mintaátlag torzítatlan és konzisztens becslése a ϑ igazi paraméternek.

Vagyis egy becslés jó, ha torzítatlan és főleg konzisztens.

Állítás: Ha $m(\vartheta) < \infty$, akkor a mintaátlag konzisztens becslése a várható értéknek:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n\to\infty]{} m(\vartheta)$$

A nagy számok törvénye miatt.

ha: $X_1, X_2, \dots, X_n \sim EXP(\vartheta)$ $\quad \vartheta \in (0, \infty)$

akkor az igazi $\vartheta = ?$

$$\frac{X_1 + \dots + X_n}{n} \to \frac{1}{\vartheta}$$

az exp. eloszlás várható értéke

$$\Downarrow$$

$$\frac{n}{X_1 + X_2 + \dots + X_n} \to \vartheta$$

ez a statisztika konzisztens becslése az exp. eloszlás paraméterének

Tény: $S_n^2 \xrightarrow[n\to\infty]{} \sigma^2(\vartheta)$ A tapasztalati szórásnégyzet konzisztens becslése a szórásnégyzetnek.

Ha van egy konkrét feladat

$\downarrow$

példa:

$$f(x) = \frac{\vartheta}{x^{\vartheta+1}} \qquad x \geq 1$$

1,53 ; 2,76 ; ...

↑ van egy 10 elemű minta (adathalmaz)

Ekkor az igazi ϑ mi?

↓

ez már becslési eljárás.

↓

Becslési eljárások

- Maximum likelihood ~~(MK)~~ (ML)
- momentum módszer
- legjobb négyzetes becslések

1, Maximum likelihood becslések

Feldobunk 10-szer egy pénzérmét $(x_1, x_2, \dots, x_{10})$. ϑ valószínűséggel fej.

$$P(\text{fej}) = \vartheta \text{, ahol} \qquad \Theta = (0,1)$$

$$0 < \vartheta < 1$$

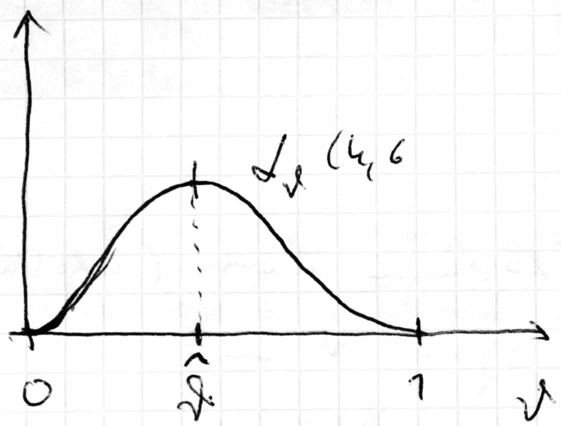
Találjuk meg ϑ igazi értékét!

10 mérés: 4 fej, 6 írás.

$$L_\vartheta(4,6) = P_\vartheta(x_1, \dots, x_{10} \text{ mintából } 4 \text{ fej}, 6 \text{ írás}) =$$

↑ likelihood - fgv

$$= \binom{10}{4} \vartheta^4 \cdot (1-\vartheta)^6$$



Azt a ϑ-t fogadjuk el, amelyik a maximum → itt a legnagyobb a valószínűsége, hogy a $(4,6)$ pontot kapjuk: ez $\hat\vartheta$. Ez lesz a maximum likelihood becslés.

$$\partial_\vartheta L(4,6) = 0$$ egyenlet megoldása

↑ deriválunk ϑ szerint

$$4\cdot\binom{10}{4}\vartheta^3\cdot(1-\vartheta)^6 - 6\cdot\binom{10}{4}\vartheta^4(1-\vartheta)^5 = 0$$

ebből $\vartheta = \frac{4}{10}$,

vagyis $\hat\vartheta = \frac{4}{10}$ (itt lesz a maximum).

Általában a max. likelihood becslés:

Adott $X_1, X_2, \dots X_n$ minta P_ϑ eloszlással, a megvalósulás $\underline{x} = (x_1, x_2, \dots, x_n)$

Ha P_ϑ diszkrét:

$$L_\vartheta(x_1, x_2, \dots, x_n) = P_\vartheta(X_1 = x_1, \dots X_n = x_n) =$$

↑ a likelihood-fgv.

$$= \prod_{i=1}^{n} P_\vartheta(X_i = x_i)$$, mert függetlenek, azonos eloszlású P_ϑ eloszlású)

Ha P_ϑ folytonos eloszlás: f_ϑ sűrűségfüggvénnyel:

$$\mathcal{L}_\vartheta(x_1,\dots,x_n) = \prod_{i=1}^{n} f_\vartheta(x_i) =$$

$$= \lim_{\varepsilon\to 0} \frac{1}{(2\varepsilon)^n} P_\vartheta\left(X_1\in[x_1-\varepsilon, x_1+\varepsilon], \dots X_n\in[x_n-\varepsilon, x_n+\varepsilon]\right)$$

melyik ϑ-ra lesz maximális. → azt a ϑ-t, amire a kimenetel a legvalószínűbb.

Ill. azt a ϑ-t vesszük, ami mellett az $x_1,\dots,x_n$ körüli kis gömbökbe esés a legvalószínűbb.

Def: Maximum likelihood becslés:

Egy $\hat{\vartheta}$ statisztika ($(X_1,\dots,X_n)$-et a fgv-e) a ϑ paraméter max. likelihood becslése, ha a $\hat{\vartheta}$ globális maximuma a $\vartheta \mapsto \mathcal{L}_\vartheta(x_1,\dots,x_n)$ $\forall x_1,\dots,x_n$-re.

Melyik paraméterérték mellett lesz a paraméterérték a max valószínűségű → ez a $\hat{\vartheta}$

A, $\hat{\vartheta}$ kiszámítása

B, Mennyire jó ez a becslés ϑ-nak?

A, $\hat{\vartheta}$ kiszámítása, ha Θ konvex és nyílt, $\mathcal{L}_\vartheta$ differenciálható ϑ-ban, akkor $\partial_\vartheta \mathcal{L}_\vartheta(\underline{x}) = 0$ (deriválttal) egyik megoldása a globális maximum.

B, A torzítatlanságról nem tudunk sokat mondani, viszont bizonyos feltételek mellett (gyenge) konzisztens

Példa: Csalódott jövedelmét olyan skálán mérjük, ahol x_1 a létminimum...

Adott P_ϑ
$$f_\vartheta(t) = \frac{\vartheta}{t^{\vartheta+1}}, \text{ ha } t \geq 1 \quad \& \text{ sűrűségfgv-nyel.}$$

$$\vartheta \in (0, \infty)$$

ezek meg vannak adva.

Adott egy 10 elemű minta: $(x_1, \dots, x_{10})$

Adjunk max. likelihood becslést ϑ-ra!

$$f_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i)$$

hol maximális?

Logaritmust vesszük, hogy könnyebb legyen deriválni

$\downarrow$ mivel $\log()$ m. n. $\uparrow$, ezért ugyanígy viselkedik.

$\downarrow$

$$l_\vartheta(\underline{x}) = \log L_\vartheta(\underline{x}) \qquad \text{ez a log likelihood-fgv.}$$

Ennek keressük a maximumát.

~~Kez~~ l_ϑ-val csak dolgozunk:

$$L_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\vartheta}{x_i^{\vartheta+1}}$$

$$l_\vartheta(x_1, \dots, x_n) = \sum_{i=1}^{n} \left(\log \vartheta - (\vartheta+1) \cdot \log x_i\right) =$$

$$= n \cdot \log \vartheta - (\vartheta+1) \cdot \sum_{i=1}^{n} \log x_i$$

Ez maximális → deriváljuk ϑ szerint:

$$\partial_\vartheta \, l_\vartheta(\underline{x}) = \frac{n}{\vartheta} - \sum_{i=1}^{n} \log x_i = 0$$

$$\Downarrow$$

$$\vartheta = \frac{n}{\sum_{i=1}^{n} \log x_i}$$

ez a max. likelihood becslés:

$$\boxed{\hat{\vartheta} = \frac{n}{\sum_{i=1}^{n} \log x_i}}$$

HF: $EXP(\lambda)$ max. likelihood becslése $(x_1, \dots, x_n)$ minta alapján:

$$L_\vartheta(x) = \prod_{i=1}^{n} \lambda \cdot e^{-\lambda x_i}$$

előadás 2010. XII. 02.

ÉZT ELŐADÁSDIÁBÓL!

Statisztika (folytatás)

- becslések — pont ~~[illegible]~~ becslések: P_ϑ, $\vartheta \in \Theta$ és az $x_1, x_2, \dots, x_n$ minta alapján akarjuk becsülni a ϑ igazi paramétert.
- hipotézisvizsgálat

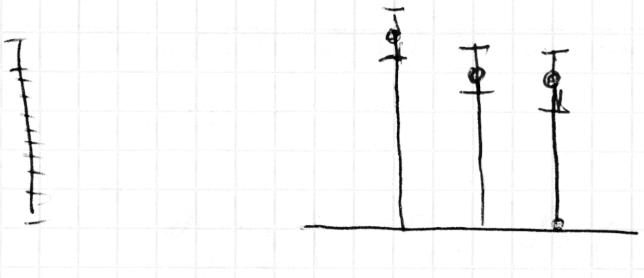
A pontbecslésben

A becslés közel van az igazihoz: $T(\underline{x})$ közel van az igazihoz.

$\underline{x} = (x_1, \dots, x_n) \longrightarrow \vartheta$ igazi paraméter.

Intervallumbecslések

Egy érték helyett egy intervallumot adunk a paraméter becslésére



A beállítás ugyanaz: adott P_ϑ

Az intervallum két végpontját kell megadni.

A $(T_1(x), T_2(x))$ statisztika párral definiált intervallum legalább $1-\varepsilon$ szintű konfidencia intervallum a $\Psi(\vartheta)$ paraméterre, ha $P_\vartheta(T_1(\underline{x}) < \Psi(\vartheta) < T_2(\underline{x})) \geq 1-\varepsilon$

$\forall \vartheta \in \Theta$

ϑ az igazi paraméter, a két végpont közé essen ε paraméterrel.

~~Jelentése: legyen $T(\vartheta) = \vartheta$, akkor~~

Ha veszünk nagyon sok adatsort:

$$\left.\begin{array}{l} x^{(1)}, \dots \longrightarrow \left(T_1(\underline{x}^1),\ T_2(\underline{x}^2),\ \dots\right) \\ \text{mérési adatsor} \\ \vdots \end{array}\right\} \text{intervallumok.}$$

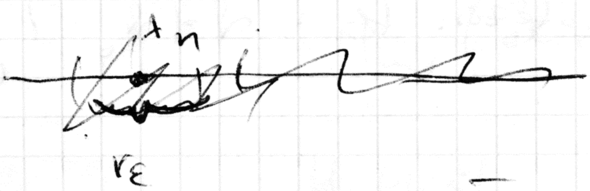
Ezekbe a mérések 95%-ban beleesik a paraméter a kívánalttól,.

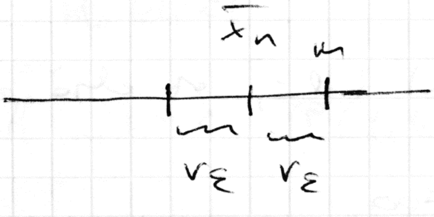
Példa: $\mathcal{N}(m, 20)$ 30 elemű minta.
Erre 95% konfidencia intervallum.

Az intervallum középére jó a mintaátlag. Előbb akarunk egy intervallumot kapni.

$$P_m(\bar{x}_n - v_\varepsilon < m < \bar{x}_n + v_\varepsilon) = 1 - \varepsilon$$

Átrendezéssel v_ε-t lehet megkapni.





$$P_m\left(-\frac{r\varepsilon}{\sigma_0}\sqrt{n} < \frac{\bar{x}_0 - m}{\sigma_0}\sqrt{n} < \frac{r\varepsilon}{\sigma_0}\sqrt{n}\right)$$

$$\frac{(x_1+x_2+\dots+x_n)-nm}{\sqrt{n}\cdot\sigma_0} \sim N(0,1)$$

normális eloszlás

így $\phi\left(\frac{r\varepsilon}{\sigma_0}\sqrt{n}\right)$

10 elemű mintát veszünk és ismételjük. Intervallumot.

Végtelen sok minta -> az esetek 95%-ában a paraméter benne lesz az intervallumban.

<u>Paraméteres próbák</u>

Van egy P_ϑ $\vartheta \in \Theta$ eloszláscsalád.

Van egy igazi paraméter, amiből legenerálódik a minta.

Van egy nullhipotézis: $H_0: \vartheta = \vartheta_0$ } el kell dönteni,
alternatív hipotézis: $H_1: \vartheta \neq \vartheta_0$ } hogy melyik igaz.

A szórást nem tudhatjuk, az előző példában azt feltettük.

Ellenhipotézis: nem 300.

A hipotézisvizsgálat filozófiája: mindenkit ártatlannak tételezünk fel.

Ha a konfidenciaintervallum tartalmazza a 300-at, akkor elfogadjuk.

95%

99% – még többet megengedünk.

Azonos elsőfajú hiba másodfajú hibája kisebb.

$$P(\text{elsőfajú hiba}) = \varepsilon = 1-(\text{szignifikancia szint})$$

$$m_0 \in \left(T_1(\underline{x}), T_2(\underline{x})\right) \Leftrightarrow H_0 \text{ elfogad.}$$

$$P_{m_0}\left(m_0 \in \left(T_1(\underline{x}), T_2(\underline{x})\right)\right) = 1-\varepsilon$$

P_{m_0} ↑ az m_0 teljesül, azaz pont H_0-nál feltételeztük.

H_0 teljesül és H_0-at elfogadjuk.

Elsőfajú hiba valószínűsége:

$$= \underbrace{P_{m_0}}_{H_0 \text{ teljesül}}(\text{elvetjük}) = P_{m_0}\left(m_0 \notin \left(T_1(\underline{x}), T_2(\underline{x})\right)\right) = \varepsilon$$

Konf. szint $= 1 - P(1.\ \text{hiba})$.

Minél nagyobb a szignifikancia-szint

H_0 nem bizonyul semmit.

Teszt-statisztikát veszünk -> ennek az értéke alapján döntünk.

Ha H_0 teljesül a teszt-statisztikának kell tudni az eloszlását.

H_0 -> az m_0 eloszlással rendelkezőből generálódott.

$$\sqrt{n}\cdot\frac{\bar{x}_n - m_0}{\sigma_0} \to 0$$

Ha H_0 nem teljesül, akkor

$$\sqrt{n}\cdot\frac{\bar{x}_n - m_0}{\sigma_0} \to \pm\infty$$

$$P_{m_0}\Big(-x < \underbrace{\text{teszt. stat.}}_{\sim N(0,1)} < +\varepsilon$$

A mintaátlag: $$\frac{x_1+\dots+x_n}{n} \sim N(m_0, \sigma)$$

Döntés: a tesztstatisztika milyen értékei mellett kell H_0-at (H_1-et) elfogadni.

tesztstatisztika

$-K < T(\underline{x}) < K \iff H_0$-at elfogadjuk.

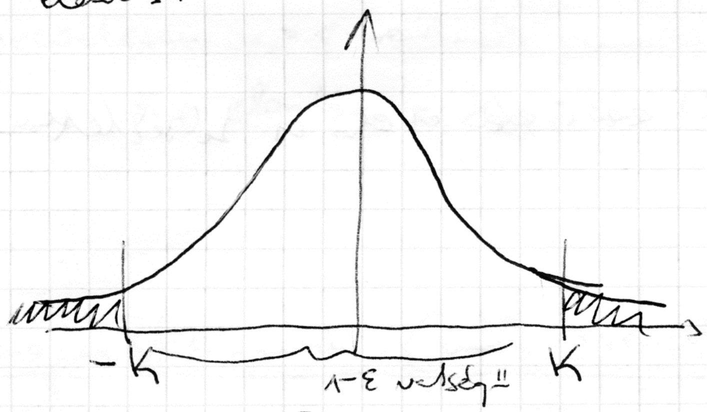
Ehhez a tesztstatisztikát kell tudni.

$H_0: m-n$

$$P_{m_0}\left(-K < T(\underline{x}) < K\right) = 1-\varepsilon$$

H_0-át elfogadjuk $(1-\varepsilon)$ szignifikancia-szinten.

t-eloszlás:



ε a kettőnek együtt

A t-eloszlásos táblázatban csak az egyiknek van.

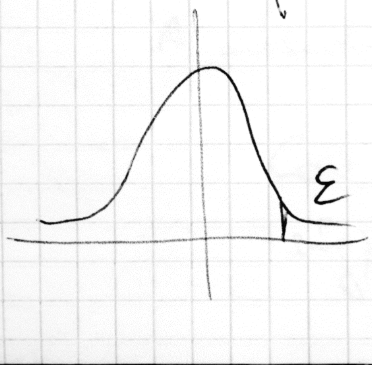
Elfogadod H_0 $(1-\varepsilon)$ szinten, ha

$$P_{m_0}(-K < T < K) = 1-\varepsilon$$

2 farok valsége 5%, ekkor 1 farokné 2,5%.

Egyoldali és kétoldali ellenhipotézis

Egyoldali hipotézis



ε ← az egész farok megvan.

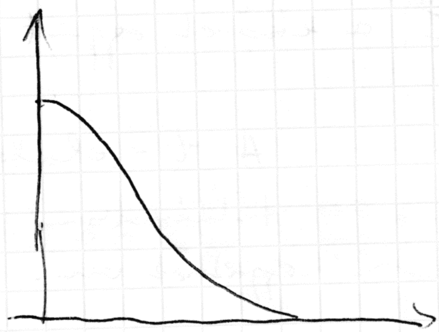
Nemparaméteres próbák

- illeszkedésvizsgálat
- homogenitásvizsgálat
- függetlenségvizsgálat

Khi-négyzet eloszlás

$U_i^{(n)} \sim BINOM(n, p)$

hányszor fordult elő az i edik elem az n kísérletben.



illeszkedésvizsgálat

. előadás 2010. XII. 07.

Stacionárius folyamatok

Veszünk egy időintervallumot, nézzük a folyamat eloszlását = veszünk egy másik időintervallumot és nézzük a folyamat eloszlását.

↓ erősen stacionárius folyamat

$(X_t, t \in T)$ erősen stacionárius, ha $\forall t_1, t_2, \dots t_n$ $\forall s$-re

$(X_{t_1}, \dots X_{t_n})$ eloszlása $= (X_{t_1+s}, X_{t_2+s}, \dots, X_{t_n+s})$ eloszlása.

speciálisan $n = 1$-re:

X_t eloszlása $= X_{t+s}$ eloszlása ($\forall$ időben ugyanazt látjuk)

Gyengén stacionárius (másodrendben stacionárius) folyamatok

Ha erősen stac, akkor a várható érték $\forall t$-ben ugyanaz. Egy intervallumban a szórásnégyzetet ugyanazok erősnél, ebből megőrzünk valamit:

legyen $(X_t, t \in T)$ $T = \mathbb{Z}$ vagy $\mathbb{R}$ gyengén stacionárius folyamat, ha

$\dots x_{-2}, x_{-1}, x_0, x_1, x_2, \dots$

$E(X_t) =$ konstans (t-től független)

↑ ezt megőrizzük az erősből elsőrendben.

A másodrendű momentumokban:

$Cov(X_t, X_s)$ csak t és s különbségétől (távolság) függ: $|t-s|$

$C_X(u) = \mathrm{cov}(X_t, X_{t+u})$ az X folyamat autokovariancia függvénye.

Ez szimmetrikus, mert csak a távolságtól függ:

$$C_X(u) = C_X(-u)$$

és $C_X(0) = \mathrm{cov}(X_t, X_t) = \sigma_X^2$ (ami $\forall$ időben ugyanaz).

~~Gyengén stac. folyamat pl.~~:

legyen

X_i, $i \in \mathbb{Z}$ független, azonos eloszlásúak.

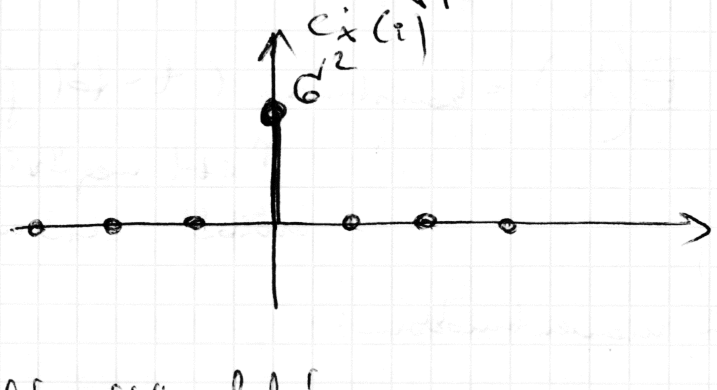
$$m = E(X_i)$$
$$\sigma^2 = D^2(X_i)$$

ennek az autokovariancia-függvénye:

$$\mathrm{cov}(X_i, X_j) = \begin{cases} 0, & \text{ha } i \neq j \end{cases}$$

ha két vletlen független, akkor 0 a cov

Az autokovariancia fgv:



~~Általánosított~~ fehérzaj: ↓

X_i, $i \in \mathbb{Z}$ $X_i \in N(0,1)$

Általánosított fehérzaj:

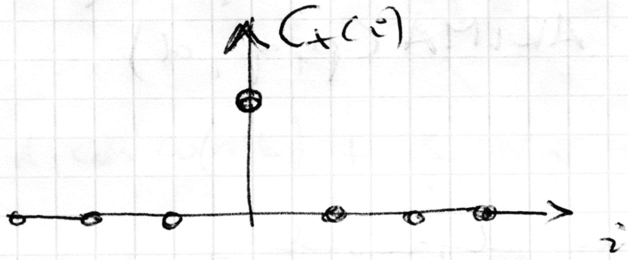
$$(\xi_i \quad i \in \mathbb{Z}) \text{ -re } \quad E(\xi_i) = 0$$

$$\text{és } COV(\xi_i, \xi_j) = 0 \text{ , ha } i \neq j.$$

például: $\xi_i = X_i X_{i+1} X_{i+2}$

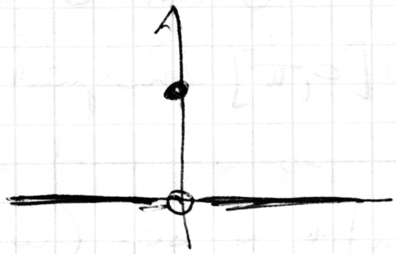
$$E(\xi_i) = \emptyset$$

Az ált. fehérzaj ~~kov~~ autokovariancia fgv-e:



Lehetne ~~ezt~~ folytonos időben és nézni.

Folytonos idejű fehérzaj:



Lehet komplex értékű folyamatokat és definiálni.

$$COV(X_t, X_s) = E(X_t - EX_t)(X_s - EX_s)$$

komplexnél bejön egy konjugálás.

$X_t \in t \in \mathbb{R}$ $\mathbb{C}$ értékű

$$COV(X_t, X_s) = E\,\overline{(X_t - EX_t)}(X_s - EX_s)$$

Stacionárius folyamatosztályok

1.) Példák

2.) - Stacionárius Markov-lánc

- Stacionárius, folytonos idejű Markov-lánc

2.) Trigonometrikus folyamatok

3.) Mozgóátlag folyamatok ($MA(q)$)

4.) Autoregresszív folyamatok ($AR(p)$)

5.) Autoregresszív és mozgóátlag folyamatok $ARMA(p,q)$

Integrált autoregresszív és mozgóátlag folyamatok

$ARIMA(p,q,d)$

2.) Trigonometrikus folyamatok

Legyen A és B korrelálatlanok $\to$ $COV(A,B)=0$.

és legyen olyan, hogy $EA = EB = 0$.

és $\sigma^2(A) = \sigma^2(B)$ ~~ma~~

Legyen egy $\omega \in [0,\pi]$ szögérték is. Ekkor

$$X_n = A \cdot \cos(\omega n) + B \cdot \sin(\omega n)$$

<u>**ÁLL**</u>: Ez az ~~X_n~~ $\ldots X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ egy stacionárius folyamat.

(A szög determinisztikus, de az amplitúdót megválaszthatjuk $\forall$ időben).

<u>Biz</u>: Kell: $E(X_n) =$ konstans

és $COV(X_n, X_{n+k})$ n-től független.

$E(X_n) =$ konstans, mert

$$EX_n = \underbrace{EA}_{\phi} \cdot \cos(\omega n) + \underbrace{EB}_{\phi} \cdot \sin(\omega n) = 0$$

ezeket így vméltük meg.

illetve

$$\mathrm{cov}(X_n, X_{n+k}) = E\left(X_n - \underbrace{EX_n}_{=\phi}\right)\left(X_{n+k} - \underbrace{EX_{n+k}}_{=\phi}\right) =$$

$D^2(X) = E(X^2) - \underbrace{E(X)^2}_{=0}$

$= E(X^2)$

$$= EX_n X_{n+k} = E\Big((A \cdot \cos \omega n + B \sin \omega n) \cdot$$

$$\cdot \left(A \cdot \cos \omega(n+k) + B \cdot \sin \omega(n+k)\right)\Big) =$$

két valváltozó van: A és B.

$$= E\left(A^2 \cdot \cos(\omega n) \cdot \cos(\omega(n+k)) + B^2 \cdot \sin(\omega n + \sin(\omega k))\right)$$

$EA^2 = EB^2 = \sigma^2$

$$= \sigma^2 \cdot \cos(\omega k) = C_X(k) = C_X(-k)$$

a koszinusz szimmetrikus

Példa: Lerögzítünk néhány véletlen amplitúdót és szöget: szlegpen $A_0, A_1, \dots, A_m$ és $B_1, B_2, \dots, B_m$ korrelálatlanok (mindenhogyan), és várható értékük ϕ. Az A_k és B_k négyzetnégyzete σ_k^2.

Lerögzítünk néhány ω-t $[0, \pi]$-ből.
$(\omega_1, \omega_2, \dots, \omega_m)$

A definiált trigonometrikus folyamat:

$$X_n = A_0 + \sum_{k=1}^{m} \left(A_k \cdot \cos(\omega_k n) + B \cdot \sin(\omega_k n)\right)$$

<u>**Áll**</u>: Az $(X_n, n \in \mathbb{Z})$ folyamat stacionárius folyamat.

~~teppen~~ $$\mathrm{COV}(X_n, X_{n+v}) = \sum_{k=1}^{m} \sigma_k^2 \cos(v \cdot \omega_k) = C_X(v)$$

$$E(X_n) = 0.$$

Ha vesszük a hányadok összegét:

$$C_X(0) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_m^2 = \sigma^2$$

Ekkor $$C_X(v) = \sigma^2 \cdot \sum_{k=1}^{\infty} \cos(v \cdot \omega_k) \cdot \frac{\sigma_k^2}{\sigma^2}$$

$\frac{\sigma_1^2}{\sigma^2}, \frac{\sigma_2^2}{\sigma^2}, \dots, \frac{\sigma_k^2}{\sigma^2}$ egy eloszlás $1, 2, \dots, k$-n

Értelmezés: $v \cdot \omega_k$ frekvencia mennyire van jelen az X_n trigonometrikus folyamatban.

$\Downarrow$

ω_k frekvencia $\frac{\sigma_k^2}{\sigma^2}$ súllyal járul hozzá a kovarianciához.

Pl. $X_t = \cos(\omega t + \phi)$, ahol $\phi \sim \mathrm{Unif}[0, 2\pi]$ egyenletes

X stacionárius folyamat

(*)

3, Mozgóátlag - folyamatok

Van egy $\xi_n, n \in \mathbb{Z}$ fehérzaj.

$$X_n = b_0 \xi_n + b_1 \cdot \xi_{n-1} + \dots + b_q \cdot \xi_{n-q}, \text{ ahol}$$

$\uparrow$ súlyozzuk

b_i-k valós számok.

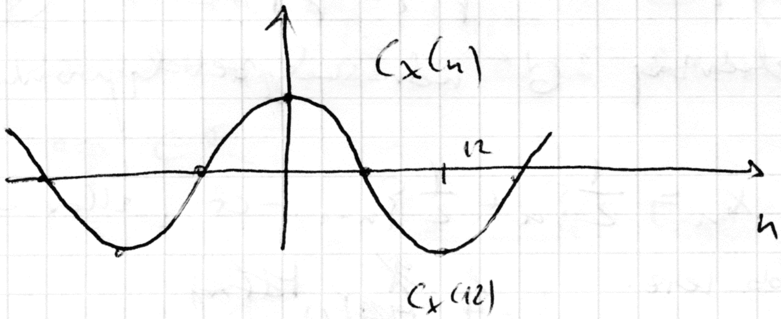
$b_0, b_1, \dots, b_q \in \mathbb{R}$, $b_q \neq 0$.

Ezt könnyű kiszámolni.

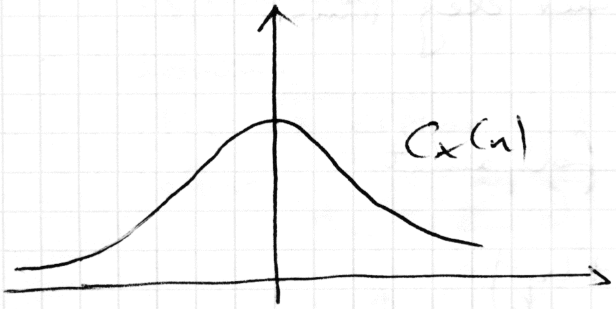
---

(4) Egy darab trigonometrikus:

$$C_X(n) = \sigma^2 \cdot \cos(\omega n)$$



Valamilyen lin. összefüggés a valváltozók között.
Mellora a lin. összefüggés a jelennel. (a távoli jövővel és múlttal is összefügg).



Ha gyorsan cseng le, akkor az összefüggés a jelenre kicsi.

X-nek hosszútávú a memóriája, ha $\sum_{n=1}^{\infty} C_X(n) = \infty$ és

↓ pl. vonater

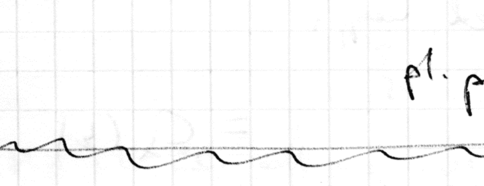
rövidtávú, ha $\sum_{n=1}^{\infty} C_X(n) < \infty$.

↓ A ~~vonatokon is ilyen~~

pl. pénzügyi folyamatok.

---

példa: $(X_n, n \in \mathbb{Z})$ fehérzaj



Visszatérve a mozgóátlaghoz:

Def: $(X_n, n \in \mathbb{Z})$ mozgóátlag folyamat (MA(q)), ha

$\exists b_0, b_1, \dots, b_q \in \mathbb{R}$, $b_0 \neq 0$ egy, hogy ~~$X_n =$~~

$$X_n = b_0 \xi_n + b_1 \xi_{n-1} + \dots + b_q \xi_{n-q},$$ ahol $(\xi_n, n \in \mathbb{Z})$

általánosított fehérzaj σ^2 szórásnégyzettel, μ várható értékkel.

Ha veszünk egy $X_n = \frac{1}{2}\xi_n + \frac{1}{2}\xi_{n-1}$-et, akkor a fehér-zaj kicsit simább lesz.

↑ MA(1)

~~$X_n =$~~

$$X_n = \frac{1}{7}\left(\xi_n + \dots + \xi_{n-6}\right) \in MA(6)$$

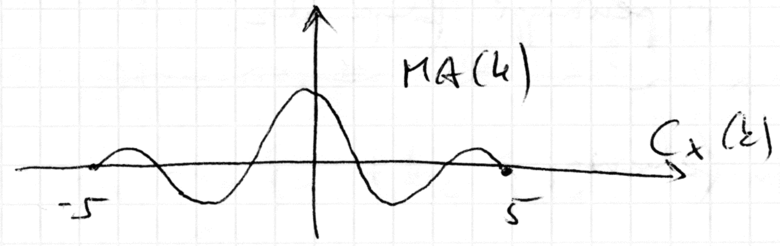
ez már elég sima.

Áll: MA(q) stacionárius folyamat

$$E(X_n) = \mu(b_0 + b_1 + \dots + b_q)$$

$$COV(X_n, X_{n+k}) = \begin{cases} 0, \text{ ha } k > q \\ \sigma^2(b_q b_{q-k} + \dots + b_k b_0), \text{ ha } k \le q \end{cases}$$

Olyan kovarianciát is lehet, hogy:



$= C_X(k)$

ha megadunk előre kovarianciákat, léteznek-e b-k.

Ha két fgtlen. mozgóátlag-folyamatot összeadunk, akkor az összeg is mozgóátlag, a múlt maximumig kell visszamenni.

4.) Autoregresszív folyamatok

Def: Adott egy $\xi_{n,m} \in \mathbb{Z}$ fehérzaj σ^2 szórásnégyzettel és 0 várható értékkel és adott $a_1, a_2, \ldots, a_p \in \mathbb{R}$.

Ha az X_n folyamat kielégíti az $X_n = a_1 X_{n-1} + a_2 X_{n-2} + \ldots + a_p X_{n-p} + \xi_n$ egyenletet, akkor X autoregresszív folyamat (AR(p)).

Áll: X_n stacionárius, ha $x^p - a_1 x^{p-1} - a_2 x^{p-2} - \ldots - a_p = 0$ egyenletnek minden gyöke a komplex egységkörön.

Ha az összes gyök az egységkörön belül van, akkor

$$X_n = \sum_{k=0}^{\infty} \delta_k \xi_{n-k}$$

$\uparrow$ $\xi_n, \xi_{n-1}, \xi_{n-2}, \ldots$ előállítja.

Ha $\forall$ gyök az egységkörön kívül van, akkor

$$X_n = \sum_{k=0}^{\infty} \delta_k \xi_{n+k}$$

$\xi_n, \xi_{n+1}, \xi_{n+2}, \ldots$ (jelen + jövő)

Ha itt is-ott is van gyök, akkor minden pillanattól függ.

Autoregresszív: a fehérzaj 0 várható értékű. A jelen csak a folyamat múltjától függ. A folyamat valahogy önmagát becsli.

ez a múltja

Példa: AR(1)

$$X_n = a \cdot X_{n-1} + \xi_n$$

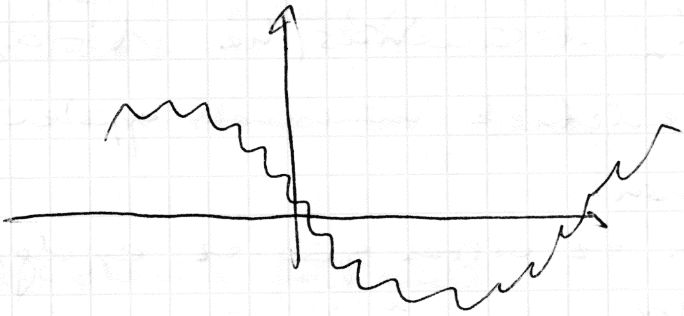
ha $-1 < a < 1$, akkor az egységkörön belül van a gyök, így végtelen mozgóátlaggal előállítható.

$x - a = 0$

$x = a$

Ha $a = \frac{1}{2}$: $X_n = \frac{1}{2} X_{n-1} + \xi_n$

valamennyit számít rajta



Ha viszont $X_n = 0{,}99 X_{n-1} + \xi_n$

jobban kiegyenlíti.

Kovarianciája és várható értéke:

$$E(X_n) = 0$$

Ugyanis $X_n = aX_{n-1} + \xi_n = a(aX_{n-2} + \xi_{n-1}) + \xi_n =$

$= a^2 X_{n-2} + a\xi_{n-1} + \xi_n = \dots =$

$= \underbrace{a^k X_{n-k}}_{} + \xi_n + a\,\xi_{n-1} + \dots + a^{k-1}\xi_{n-k}$.

~~a<1, így ezt véve = vethető értéket:~~

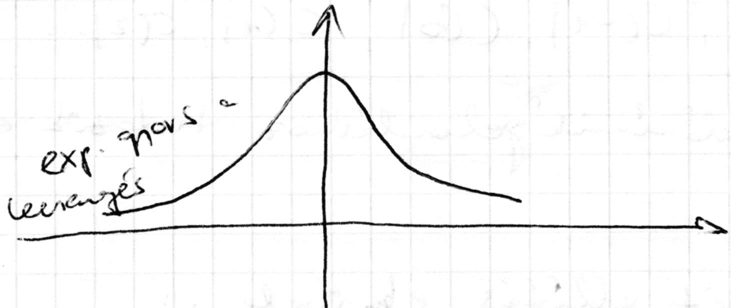
ő a hiba.

$$E\left(X_n - \left(\xi_n + \dots + a^{k-1}\xi_{n-k}\right)\right)^2 = a^{2k}\underbrace{E X_{n-k}^2}_{\sigma^2} \to 0$$

$a^{2k} \to 0.$

Vagyis $X_n = \xi_n + a\,\xi_{n-1} + a^2\xi_{n-2} + \dots$

$$\mathrm{cov}(X_n, X_{n+k}) = \dots = |a|^k\left(\frac{\sigma^2}{1-|a|}\right)$$



Ha $a>1$ egy AR(1)-ben:

$$X_n = \xi_n + \frac{1}{a}\,\xi_{n+1} + \frac{1}{a^2}\,\xi_{n+2} + \dots$$

ha nincs gyök az egységkörön, akkor stacionárius.

$X_n = X_{n-1} + \xi_n$ nem stacionárius folyamat, ezt $a=1$. (rajta van az egységkörön)

Spektrum

$$X_n = A_0 + \sum_{k=1}^{m}\left(A_k\cdot\cos(\omega_k n) + B_k\cdot\sin(\omega_k\cdot n)\right)$$

$\omega_k \in [0,\pi]$

legyen $A_0 = 0$

$$\operatorname{cov}(X_n, X_{n+v}) = \cancel{\sigma_0^2} + \sum_{k=1}^{m} \sigma_k^2 \cdot \cos(\omega_k \cdot v) =$$

$$= \sigma^2 \left(\frac{\cancel{\sigma_0^2}}{\sigma^2} + \sum_{k=1}^{m} \frac{\sigma_k^2}{\sigma^2} \cdot \cos(\omega_k \cdot v) \right)$$

ω_k frekv. ekkora súllyal járul hozzá a kovarianciához.

~~Egy diszkrét eloszlás esetén:~~

<u>Tétel</u> (Spektrálmérték létezése):

$C(k) \quad k \in \mathbb{Z}$

$\ldots C(-2), C(-1), C(0), C(1), C(2) \ldots$ adott fgv.

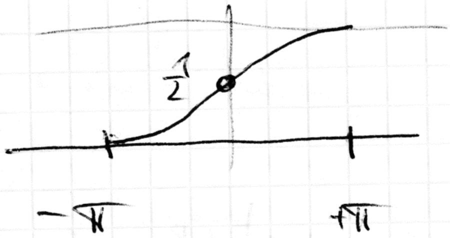
Milyen valószínűségeloszláshoz tartozik egy kovariancia?

A következő két állítás ekvivalens:

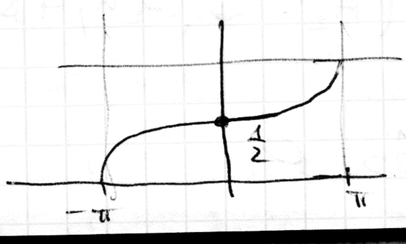
A) $C(k)$ egy valós értékű, 0 várható értékű, 1 szórású stacionárius folyamat autokovariancia fgv-e ~~pontosan akkor, ha~~

B) Létezik egy $[-\pi, \pi]$ szimmetrikus ~~fgv~~ eloszlásfgv-e (F)

pl:



vagy



úgy is hogy $C(k) = \int_{-\pi}^{\pi} \cos(k\omega)\, dF(\omega)$, ha

F folytonos eloszlás f sűrűségfgv.-e

$dF(\omega) = f(\omega)\, d\omega$

$\int_{-\pi}^{\pi} e^{ik\omega}\, dF(\omega)$: Fourier.

. előadás 2010. XII. 09.

Stacionárius folyamatok

Másodrendben stacionaritást nézünk.

kovarianciák csak az időkülönbségek absz. értékétől függenek.

Stacionárius folyamatok spektruma

A stac. folyamatot jellemzi a várható érték $EX_t = m$ és a kovariancia-fgv: $C_X(k) = \mathrm{cov}(X_t, X_{t+k})$, $k \in \mathbb{Z}$; valamilyen frekvenciák milyen valószínűséggel vesznek részt a kovarianciák előállításában.

Tétel: Legyen $C(k)$, $k \in \mathbb{Z}$ adott fgv. A következő két állítás ekvivalens.

(A) A $C(k)$, $k \in \mathbb{Z}$ egy 0 várható értékű, 1 szórású ($C(0) = 1$) stac. folyamat kovariancia fgv-e.

(B) ~~Ha~~ A $[-\pi, \pi]$-n létezik egy szimmetrikus valószínűség-eloszlás F eloszlásfgv-nyel úgy, hogy

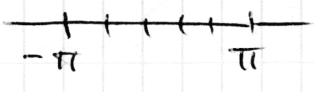
$$C(k) = \int_{-\pi}^{\pi} \cos(k\omega)\, dF(\omega) = \int_{-\pi}^{\pi} e^{ik\omega}\, dF(\omega)$$

Mivel F szimmetrikus, ezért ez valamilyen Fourier-transzformált

$(e^{ix} = \cos x + i \cdot \sin x)$

F felépülhet úgy, hogy folytonos növekvő, ill. ugrásos is lehet.

Ha F ugrásos:



Ha ugrásos, az pont egy trigonometrikus folyamat előállítása:

$$X_n = A_0 + \sum_{k=1}^{n} \left(A_k \cos(\omega_k n) + B_k \sin(\omega_k n) \right)$$

Az ugrások: $\omega_1, \omega_2, \dots, \omega_m \in [0, \pi]$ vannak.

Mivel szimmetrikus, ezért:

$$-\omega_1, -\omega_2, \dots, -\omega_m \in [-\pi, 0]$$

Az ugrás nagysága pedig a következő van képviselve:

$$C_X(v) = C(0) \cdot \sum_{k=1}^{m} \frac{\sigma_k^2}{C(0)} \cos(\omega_k v)$$

ezek az ugrás-nagyságok.

~~Ha adott F, akkor~~

F a folyamat spektrálmértéke.

Láttuk: ha F ugrásos, akkor tudunk generálni olyan stacionárius folyamatot, melynek spektrálmértéke F.

Ha F nem ugrásos:

↓ ha ugrásos, akkor csak bizonyos frekvenciák vannak kiemelve, ha nem ugrásos, akkor elég sokféle frekvencia van.

Áll: Ha $C(0) + 2 \sum_{k=1}^{\infty} |C(k)| < \infty$, akkor az F-nek létezik sűrűségfüggvénye.

(a kovarianciák összegezhetők)

↓ inverz Fourier

$$C(k) = \int_{-\pi}^{\pi} \cos(\omega k)\, dF(\omega) = \{$$

$-\pi \le \omega \le \pi$

$$= \int_{-\pi}^{\pi} \cos(\omega k)\, f(\omega)\, d\omega$$

↓ ez a spektrum (spektrál sűrűségfgv)

$f(\omega)$ kiszámítása:

Adott $C(k)$, $k \in \mathbb{Z}$

$$f_C(\omega) = \frac{1}{2\pi} C(0) + \frac{1}{\pi} \sum_{k=1}^{\infty} C(k) \cdot \cos(k\omega)$$

↑

C kovariancia-struktúrához tartozó spektrál sűrűségfgv.

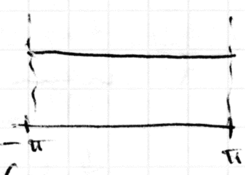
inverz Fourier: $f_C(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{ik\omega} C(k)$

<u>Néhány spektrum (példák)</u>

- fehérzaj spektruma

Ha tudjuk a kovarianciát, ki tudjuk számolni a spektrumot.

Fehérzaj: σ^2 szórásnégyzettel.



$C(0) = \sigma^2$, de $C(k) = 0$ $k \neq 0$-ra.

Ekkor $f_C(\omega) = \frac{\sigma^2}{2\pi}$ $\omega \in [-\pi, \pi]$

Ugyanis $\forall$ frekvencia előfordulása konstans.

→ ~~Példa~~

– MA(1)

$\xi_n,\ n \in \mathbb{Z}$ σ^2 szórású, 0 várható értékű fehérzaj

Ekkor pl. $X_n = \xi_n - b\,\xi_{n-1}$

Ha q magasságú a mult, 0-ban van valamilyen szórás, előre és hátra is van kovariancia.

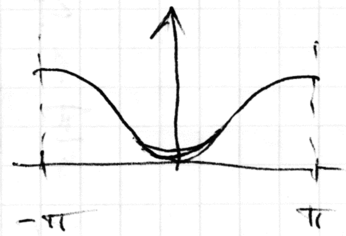
Most 1 távolság van, így

$C_X(0) = \sigma^2(1+b^2)$ az X_n szórásnégyzete.

$C_X(1) = C_X(-1) = -b\sigma^2$

egyébként: $C_X(k) = 0$

Vagyis $f_X(\omega) = \dfrac{\sigma^2(1+b^2)}{2\pi} - \dfrac{1}{\pi} b\sigma^2 \cos(\omega)$



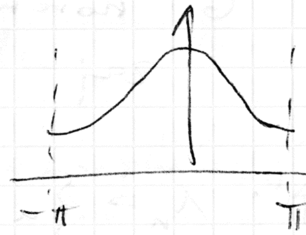
– AR(1)

$X_n = aX_{n-1} + \xi_n$

$C_X(k) = |a|^k \cdot \dfrac{\sigma^2}{1-|a|}$

kovariancia fv. a k helyen

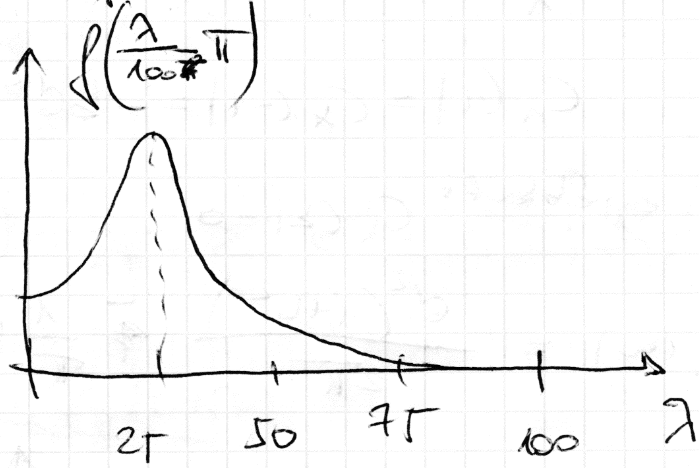
$$f_X(\omega) = \dots = \frac{\sigma^2}{2\pi\left(1-2a\cos(\omega)+a^2\right)}$$



Kialakulhatnak periódusok: pl. AR(4) és

$$X_n = 1{,}4X_{n-1} - 1{,}1X_{n-2} + 0{,}4X_{n-3} - 0{,}1X_{n-4} + \xi_n$$

ennek a spektruma:



$\uparrow$ $\frac{\pi}{4}$

$\uparrow$ a π-vel a [illegible] van feltüntetve

egy ciklus hossza így 8.

Lineáris szűrők

Van egy (gyengén) stacionárius folyamatunk: $(X_n, n \in \mathbb{Z})$

Legyen $(a_k, k \in \mathbb{Z})$ valós számok úgy, hogy

$$\sum_{k=-\infty}^{\infty} a_k^2 < \infty.$$

Ekkor $Y_n = \sum_{k=-\infty}^{\infty} a_k X_{n-k}$ az eredeti folyamat (X) a-val mint szűrője.

Jelben: $Y = a * X$

$\uparrow$ Y a-val szűri X-et.

Tetsz. időben vesszük, a jelet a_0-lal vesszük figyelembe, az 1 múltat a_1-gyel, a múltat a_{-1}, stb. Ezt lehet általánosítani folytonosra és

$(X_t,\ t \in \mathbb{R})$ stac. folyamat,

a szűrő egy valós fgv: $h(t)$, valós,

$$\int_{-\infty}^{\infty} h^2(t)\,dt < \infty$$

$$Y_t = \int_{-\infty}^{\infty} h(s) X_{t-s}\,ds$$

Miért lineáris és miért szűrő?

Lineáris, mert ha van egy a szűrő és egy b szűrő:

$a = (a_k,\ k \in \mathbb{Z})$

$b = (b_k,\ k \in \mathbb{Z})$

és ugyanazt a folyamatot szűrjük:

$$(\alpha a + \beta b) * X = \alpha a * X + \beta b * X$$

$\alpha, \beta \in \mathbb{R}$

ugyanis egy sorozatelem: $(\alpha a_k + \beta b_k)$

Egy lineáris szűrő hogy alakítja át a spektrumot?

$~~a_t~~$ $Y = a * X$

$C_X(k)$, $k \in \mathbb{Z}$ ismert

$C_Y(k)$, $k \in \mathbb{Z}$?

$\downarrow$ ki lehet számolni, hogy mi lesz a spektrum.

$f_Y(\omega) = ?$

$$C_Y(v) = \mathrm{cov}(Y_n, Y_{n+v}) =$$

$$= \mathrm{cov}\left(\sum_{k=-\infty}^{\infty} a_k X_{n-k}, \sum_{k=-\infty}^{\infty} a_k X_{n+v-k}\right) =$$

tudjuk, hogy $\mathrm{cov}(X_\alpha, X_\beta) = C_X(\alpha - \beta)$

rendezéssel

$$= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} a_r a_s \mathrm{cov}(X_{n-r}, X_{n+v-s}) =$$

$$= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} a_r a_s C_X(v-s+r)$$

úgy tűnik, mintha valami konvolúció lenne.

$$f_Y(\omega) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{ivw} \cdot C_Y(v) =$$

$$= \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{iv\omega} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} a_r a_s C_X(v-s+r) =$$

$$e^{iv\omega} = e^{i(v-s+r)\omega} \cdot e^{is\omega} \cdot e^{-ir\omega}$$

$$= \underbrace{\sum_{r=-\infty}^{\infty} e^{-ir\omega} a_r}_{\overline{A(\omega)}} \underbrace{\sum_{s=-\infty}^{\infty} e^{is\omega} a_s}_{A(\omega)} \underbrace{\frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{i(v-s+r)\omega} C_X(v-s+r)}_{\underbrace{\frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik\omega} C_X(k)}_{f_X(\omega)}} =$$

$f_X(\omega)$ spektrálsűrűség az inverz Fourier tenent.

$$= f_X(\omega) \cdot |A(\omega)|^2$$

Legyen X egy stacionárius folyamat, $a = (a_k, k \in \mathbb{Z})$, $\sum |a_k|^2 < \infty$

$b = (b_k, k \in \mathbb{Z})$, $\sum |b_k|^2 < \infty$

$c = (c_k, k \in \mathbb{Z})$, $\sum |c_k|^2 < \infty$

Először megszűrjük a-val, majd b-vel, majd c-vel.

$Y = c * b * a * X$, ennek spektruma:

$$f_Y(\omega) = f_X(\omega) \cdot |A(\omega)|^2 \cdot |B(\omega)|^2 \cdot |C(\omega)|^2,$$

ahol $A(\omega) = \sum_{k=-\infty}^{\infty} e^{ik\omega} a_k$

$B(\omega)$ és $C(\omega)$ hasonlóan.

ARMA(p, q) spektruma

ARMA(p, q):

$$X_n = a_1 X_{n-1} + \dots + a_p X_{n-p} + b_0 \xi_n + b_1 \xi_{n-1} + \dots + b_q \xi_{n-q}$$, ahol ξ egy σ^2 szórásnégyzetű fehérzaj.

Ekkor
$$\underbrace{X_n - a_1 X_{n-1} - a_2 X_{n-2} - \dots - a_p X_{n-p}}_{Y \text{ (ő lesz a szűrő)}} =$$

$Y = a' * X$

$$f_Y(\omega) = f_X(\omega)\, |A(\omega)|^2,$$

ahol $A(\omega) = 1 - \sum_{k=1}^{p} e^{ik\omega} a_k$

Tudjuk még azt is, hogy

$Y = b * \xi$

$$Y_n = b_0 \xi_n + b_1 \xi_{n-1} + \dots + b_q \xi_{n-q}$$

és $f_Y(\omega) = f_\xi(\omega) \cdot |B(\omega)|^2$,

ahol $B(\omega) = \sum_{k=0}^{q} e^{ik\omega} b_k$

Mivel a két $f_Y(\omega)$ ugyanaz, ezért

$$f_X(\omega)\, |A(\omega)|^2 = \underbrace{f_\xi(\omega)}_{\frac{\sigma^2}{2\pi} \text{ a } [-\pi, \pi]\text{-n.}} \cdot |B(\omega)|^2 = \frac{\sigma^2}{2\pi} |B(\omega)|^2$$

Vagyis

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{|b_0 + e^{i\omega} b_1 + \ldots + e^{iq\omega} b_q|^2}{|1 - e^{i\omega} a_1 - \ldots - e^{ip\omega} a_p|^2}$$

1. Példa:

Van egy $(X_n, n \in \mathbb{Z})$ stacionárius folyamatunk:

$$Y_n = \frac{1}{2} X_{n+12} + X_n + \frac{1}{2} X_{n-12}$$

Tfh $f_X(\omega)$ ismert

$$f_Y(\omega) = ? = f_X(\omega) \, |A(\omega)|^2$$

$$A(\omega) = 1 + \frac{1}{2} e^{-12i\omega} + \frac{1}{2} e^{12i\omega} = 1 + \cos(12\omega)$$

$$A\left(\frac{\pi}{12}\right) = 0$$

vagyis kiszűri a 12-es periódusokat a szűrő. (trend kiszűrése)

2. példa

$(X_n, n \in \mathbb{Z})$ stacionárius folyamat $f_X(\omega)$ spektrál-sűrűséggel.

$$Y_n = X_n - X_{n+1}$$

$a_0 = 1 \quad a_{-1} = -1 \quad a_k = 0$ különben.

$$f_Y(\omega) = f_X(\omega) \, \underbrace{|1 - e^{-i\omega}|^2}_{|1 - e^{i\omega}|^2}$$

$|e^{i\omega}|^2$-tel monoton és egy $|-1|^2$-tel és

Alkalmazzuk a műveletet r-szer:

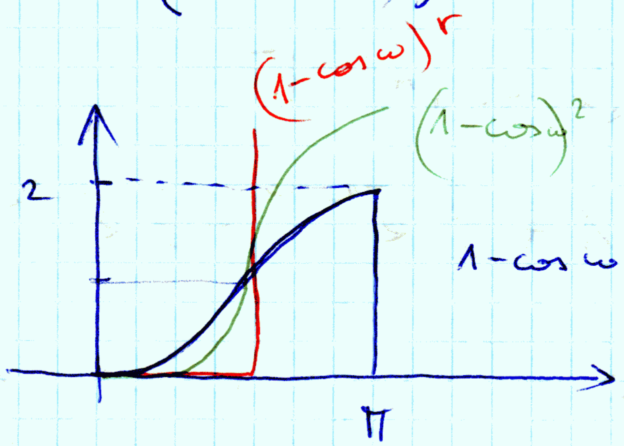
$$Y^{(r)} = \underbrace{a * a * \dots * a}_{r} * X$$

így

$$f^{r}(\omega) = f_X(\omega)\,\left|1-e^{i\omega}\right|^{2r}$$

$$\left|1-e^{i\omega}\right|^{2r} = \left|e^{i\frac{1}{2}\omega} - e^{-i\frac{1}{2}\omega}\right|^{2r} = 2^{2r}\left[\sin\left(\tfrac{1}{2}\omega\right)\right]^{2r} =$$

$$= \dots = 2^{r}\left(1-\cos(\omega)\right)^{r}$$



A magas frekvenciákat kiemeli

Ha $Z_n = X_n + X_{n+1}$

akkor $b_0 = 1$ $b_{-1} = 1$

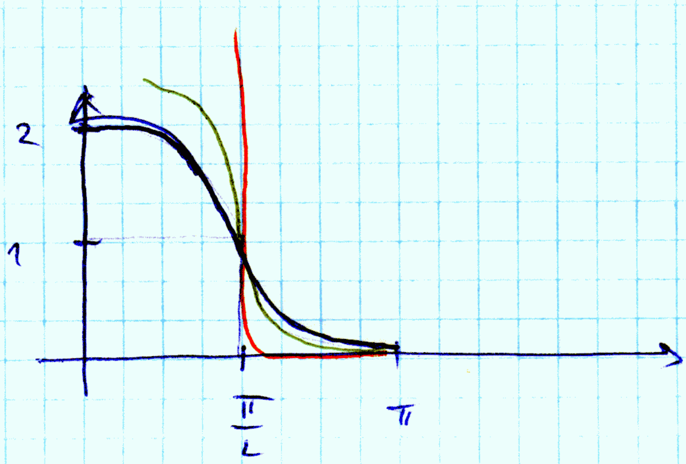
legyen $Z^{(r)} = \underbrace{b * b * \dots * b}_{r} * X$

$\cancel{f_Z(\omega)}$ $f_{Z^{(r)}}(\omega) = f_X(\omega)\left|B(\omega)\right|^{2r}$

ahol $\left|B(\omega)\right|^{2} = \left|1+e^{i\omega}\right|^{2\cancel{r}}$

Hasonló átalakításokkal, mint az előbb.

$$|B(\omega)|^{2r} = \dots = 2^r (1+\cos(\omega))^r$$



Kiemeli az alacsony frekvenciákat.

Ebből a témából elmélet lesz.