7-10-2021

Στοχαστικές μέθοδοι
Κουτσογιάννης
Μάθημα 1ο

→ Πάμε στην Όλγα κτήριο Σαντορίνη βιβλίο ΕΜΠ 2021
→ Βιβλίο Παπαδόπουλου και Κουτσογιάννη με VPN online
→ ασκήσεις που θα λύνουμε 50%
→ προφήτης ≠ στοχαστής
↳ πρόβλεψη για το μέλλον, χρησιμοποιώντας πράγματα που έγιναν στο παρελθόν.
→ για όσα πράγματα δεν έχουμε ακρίβεια, αλλά θέλουμε να τα προσεγγίσουμε δουλεύουμε στοχαστικά.

Μοντέλο:
Κάτι που αντιπροσωπεύει κάτι άλλο, δίνοντας έμφαση σε ορισμένα από τα χαρακτηριστικά του, όχι όλα.

Προσομοίωση:
Τεχνική μίμησης ενός πραγματικού συστήματος, όπως αυτό εξελίσσεται στον χρόνο.

Παράδειγμα μαθητή:
Η ετήσια εισροή I ενός ταμιευτήρα είναι κάθε χρόνο σταθερή, ίση με 10 μονάδες και η εκροή Q είναι αύξουσα συνάρτηση του αποθέματος S της μορφής: $Q(S)=0{,}2\cdot e^{0{,}3\cdot S}$

t	I	τυχαίο S	Q
0		11	
1	10	$S_0+I_1-Q_1$	$0{,}2\cdot e^{0{,}3\cdot 11}$
2	10	⋮	⋮

→ Στην αρχή μπορούμε να προβλέψουμε, μετά χάνουμε την ακρίβεια.
→ Τυχαίο λέμε το απρόβλεπτο.
Αν γράψω =RAND() και πατήσω F9 το ανανεώνει.
→ mod = υπόλοιπο

→ Εξίσωση κύκλου: $x^2+y^2=\rho^2$

→ Εμβαδόν τεταρτοκυκλίου: $\frac{\pi\rho^2}{4}$

→ s = διαστατικότητα.

→ πρέπει να λαμβάνουμε υπόψιν την αβεβαιότητα στα έργα μας.

↓ την μειώνουμε ή βρίσκουμε τρόπο να μην μας επηρεάζουν τόσο πολύ.

Αβεβαιότητα → Στοχαστικές (τυχηματικές) → aleatory πχ φυσικές αβεβαιότητες, υπάρχουν εξ' ορισμού.
→ Επιστημικές
→ Στην λήψη απόφασης

→ τα όρια είναι δυσδιάκριτα μεταξύ των κατηγοριών (μεταβάλλονται στον χρόνο ανάλογα με την γνώση)

1) η κβαντική θέση των ηλεκτρονίων στο άτομο. Αυτές δεν μπορούν να μειωθούν ακόμα και αν αυξηθεί η γνώση μας για το φαινόμενο.

2) σχετίζονται με το έλλειμμα στην γνώση μας πάνω στο φυσικό φαινόμενο. Δεν ξέρουμε να εξηγήσουμε την συμπεριφορά. πχ αβεβαιότητες στο μοντέλο. (μειώνεται καθώς αυξάνεται η γνώση ή η πληροφορία μας)

input ⇒ $\boxed{M}$ ⇒ output

1η αβεβαιότητα (στο τι βάζω)

2η αβεβαιότητα (στο πόσο καλά έχω φτιάξει το μοντέλο μου)

Επίσης από διαφορετικά δοκίμια θα πάρω διαφορετικές τιμές ⇒ parameter uncertainty.
↳ όσο πιο πολύ πληροφορία έχω τόσο καλύτερα μπορώ να το περιγράψω.

Μπορώ να έχω και τέτοια παραδείγματα στο χρόνο, εκτός από το χώρο.

3) πχ το είδος της θεμελίωσης, ή πως θα διαστασιολογήσουμε μια διατομή. Το ρίσκο δεν πρέπει να 'ναι δραματικό προς την πλευρά της αβεβαιότητας.

→ Η ανάλυση των αβεβαιοτήτων αναγνωρίζει τις πηγές της αβεβαιότητας και τις ποσοτικοποιεί.

↓ με μαθηματικό μοντέλο περιγραφής. Διαδίδει την αβεβαιότητα μέσω του μοντέλου στο αποτέλεσμα.

→ Ανάλογα με το τι θέλουμε να εκτιμήσουμε διαλέγουμε και πόσες περιπτώσεις θα πάρουμε

→ Το πόσο σημαντικό είναι το έργο θα μας κάνει να βάλουμε το όριο αστοχίας.

→ Οι συνδυασμοί δεν είναι άπειροι, αλλά τυχηματικοί.

→ Το τυχαίο πείραμα είναι το περίγραμμα του μαθήματος.

→ προσομοίωση Mode Carlo ⇒ Λύνω πολλές φορές το μοντέλο για να φτιάξουμε την καμπύλη απόκρισης ή συμπεριφοράς του μοντέλου (να βρούμε μεγέθη που διακυμαίνονται) και στην συνέχεια παίρνουμε αποφάσεις.

για διάφορες παραμέτρους εισόδου.

14-10-2021

Στοχαστικές μέθοδοι   theano_any@hotmail.com

Ηλιοπούλου, Κουτσογιάννης.

Μάθημα 2ο

→ Στην σελίδα του Κουτσογιάννη βάζουμε: more/ και μας βγάζει στις ασκήσεις.

→ Άσκηση 1_2020

1)

t	I (εισροή)	S
0	10	= 5+10·rand() (τυχαία τιμή στο (5,15), 0-1 τιμή) με F9 αλλάζει συνέχεια
1	10	= $S_0 + I_0 - 0{,}2\exp(0{,}3\cdot S)$ αι μετά φτιάχνω το ισοζύγιο
2	⋮	
3		
⋮		
100		

1α)

t	I	S	S'
0	10	=5+10rand()	= $S_0 + 0{,}0001\cdot S_0$
1	10		εδώ ίδια με αριστερά
2	⋮		μόνο την 1η αλλάζουμε
⋮	⋮		

→ μετά από κάποιο χρόνο έχουμε σημαντικές αποκλίσεις με μια μικρή διαφορά στις αρχικές συνθήκες.

→ μπορώ να κάνω DS=S-S' και να βρω από πού και μετά έχω αποκλίσεις.

1β) σε ένα κελί =COUNTIF(S,">5")/COUN(S)

→ αν αλλάξουμε κάτι η συχνότητα επηρεάζεται.

→ Προσπαθούμε να βγάλουμε μια στατιστική συμπεριφορά. Φτιάχνω 20 στήλες σαν την S, δηλαδή S_1, S_2, S_3.... για να μπορώ να μιλήσω με πιθανότητες

→ και σε κάθε μια από αυτές υπολογίζω συχνότητα.

→ για να βρω την στατιστική της συμπεριφορά mean

τυπ.αποκλ = STDEV.S(....)

→ για το 95%, τις βάζω κάθετα, κάνω sort ή small (ταξινόμηση) (εντολή) και βγάζω μόνο το 95%, το 5% των 20 τιμών είναι 1 τιμή και μπορώ να πετάξω την μεγαλύτερη τιμή.

2

1γ) Μήκος προσομοίωσης για 99% ακρίβεια
C_v = συντελεστής μεταβλητότητας.
c = σφάλμα = 0,01
∂ = 5

ποσοστημόριο κανονικής κατανομής
πχ για 95% όριο εμπιστοσύνης είναι 1,96.

$$\Rightarrow n = \left(z_{(1+\gamma)/2} \cdot C_v / c\right)^2$$

$$n = \left(1{,}96 \cdot \frac{\text{τυπ.απ.}}{\text{mean}} / 0{,}01\right)^2 \simeq 11 \text{ ή } 12$$
επαναλήψεις

→ θέλω λίγες επαναλήψεις, γιατί έχω μικρό συντ. μεταβλητότητας

1δ) Θα το βρούμε από τις διαφάνειες.

Τρόποι εκτίμησης μέσης τιμής των ροπών => Εκτίμηση της πραγματικής ροπής

2) Παίρνω πχ την χρονοσειρά των S από πριν και την βάζω σε ένα άλλο παράθυρο
→ οι 4 κεντρικές ροπές = μέση τιμή (average) mean (1), stdev (τυπική απόκλιση) (2), skewness (ασσυμετρία) (3) skew, κύρτωση (4) = kurt + 3

η κανονική κατανομή έχει μηδενική ασσυμετρία

για να οριστεί σωστά

→ για την ασσυμετρία, εκεί που είναι η ουρά μας δείχνει αν είναι θετικά ή αρνητικά ασύμμετρη.
→ όσες κατανομές είναι ψηλές (pic) έχουμε μεγάλη κύρτωση, αλλιώς είναι flat

2β) σπίτι

f_x = pdf και $F(x)$ = συνάρτηση κατανομής

2γ) Για Gauss και Gamma. (πιθανοφάνεια) μέγιστη

Πιθανοφάνεια: Έχοντας υπόθεση μια κατανομή πχ κανονική και ξέροντας το δείγμα να βρω ποιές είναι οι παράμετροι αυτής της κατανομής (οι βέλτιστες) θα είναι αυτές που μεγιστοποιούν την συνάρτηση πιθανοφάνειας: για δεδομένες τιμές δείγματος, το γινόμενο των συναρτήσεων κατανομής

(ΣΟΣ) → Θέλω ανεξάρτητα δεδομένα

Συνάρτηση πιθανοφάνειας:
Ορισμός: Το γινόμενο των συναρτήσεων πυκνότητας πιθανότητας του x (f_x) με δεδομένες τιμές παραμέτρων, δηλαδή για ένα δεδομένο m και s της κανονικής ποιό είναι το γινόμενο των συναρτήσεων κατανομής;
→ Στην ουσία ψάχνοντας την μέγιστη τιμή αυτού του γινομένου, ψάχνω την μέγιστη τιμή της από κοινού πιθανότητας να παρατηρήσω όλες αυτές τις τιμές.

→ οι παράμετροι της κανονικής κατανομής είναι το m και το s

Εστω $m = 10$ (μ.τ)

$s = 4$ (τ.α)

$\ln fx = \ln(\text{norm.DIST}(\text{τιμές}, m, s, \text{FALSE}))$

FALSE → pdf

→ βρίσκω το άθροισμα των log. ⇒ βρίσκω sum των ln

✓ Μέθοδος μέγιστης πιθανοφάνειας

να αλλάξω το 10 και το 4 ωστε να γίνει max το sum

→ Βάζω τον solver : το sum , να γίνει max, By changing m,s και solve.

→ μου δίνει κοντινές τιμές με τις κλασικές ροπές.

→ στην κανονική κατανομή : ασυμετρία = 0

κύρτωση = 3.

→ το δοκιμάζαμε και για την gamma (αλλά έχει έξτρα παραμέτρους)

~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~

→ η έννοια της πιθανότητας είναι παντού γύρω μας (στην καθημερινότητα μας)

→ το 1933 θεμελιώθηκε η θεωρία πιθανοτήτων, μοντέρνα θεωρία που αναπτύχθηκε για να καλύψει νέες εξελίξεις στην επιστήμη.

→ olga@itia.ntua.gr , για να πάρουμε το βιβλίο.

→ παίρνονται πολλές αποφάσεις με βάση πιθανότητες.

→ παλιά οι πιθανότητες είχαν ορισμούς της πλάκας (πριν το 1933)

→ Ο kolmogorov μίλησε για τα αξιώματα.

Ω → Β (βρέχει), Ξ (ξηρή)

Σ (υποσύνολα) → Β, Ξ, Β,Ξ = Ω, Φ

Φ: και να βρέχει και να μην βρέχει ⇒ αδύνατο γεγονός

→ το κενό είναι υποσύνολο κάθε συνόλου

→ και ξεκινάω να βάζω τιμές ⇒ $B = 0{,}1$, $Ξ = 0{,}9$, $B,Ξ = 1$, $Φ = 0$ } Αλλά ο kolmogorov τα είπε πιο επιστημονικά και γενικά.

→ σε σύνολα αντιστοιχούμε αριθμούς. με $P(A) \geq 0$ (αξίωμα) , $P(Ω) = 1$ (αξίωμα)

$AB = Φ \Rightarrow P(A+B) = P(A) + P(B)$ ασυμβίβαστα (αξίωμα) → κανονικά έχει άλλο 1

→ όρισε και την τυχαία μεταβλητή ⇒ μας βολεύει τα σύνολα να τα γράφουμε και αυτά με αριθμούς πχ το $B \to 1$ ή $\Xi \to 0$

→ τυχαία μεταβλητή ή στοχαστική: $\underline{X}$ θα έχει ταυτόχρονα και την τιμή 0 και 1

Παράδειγμα:

Έχουμε 1 ζάρι που το ρίχνουμε 2 φορές με $\underline{X}$ και $\underline{y}$. Ποιά η πιθανότητα:

1) $\underline{X} < \underline{y}$: 36 (6x6) όλες οι περιπτώσεις αφαιρούμε τις ίσες που είναι 6 ⇒ 30

και επειδή είναι 2 ζάρια ⇒ $\frac{30}{2} = 15$ ($1+2+3+4+5$), αλλιώς $\underline{X} < \underline{y}$ θα ήταν 21 ($1+2+3+4+5+6$)

2) $\underline{X} < y$: ($\underline{X}$ → κανονική ή τυπική μεταβλητή (είναι νούμερο πλέον))

για $y=1 \to 0$ $\quad P = 0/36$

για $y=2 \to 6$ $\quad P = 6/36$ $\quad y = \frac{1}{6}x - \frac{1}{6}$

για $y=3 \to 12$ $\quad \vdots$

για $y=4 \to 18$

για $y=5 \to 24$

για $y=6 \to 30$

3) $X < y$: Δεν έχει νόημα να μιλάμε για πιθανότητα, πρέπει η μια απ' τις 2 να είναι τυχαία. (δεν στέκει αλλιώς)

①

21-10-2021

Στοχαστικές Μέθοδοι

Κατσογιάννης, Άννυ

Μάθημα 3ο

Άσκηση 1 συνέχεια:

3) Ζήτηση = σταθερή = 5 + 5·rand( )

t	I	S
0	από το 1 ξεκινάω =norminv(rand(),6,1)	=5·rand()
1	⋮ το τραβάω	=min(max($S_0+I_1-\$Z$,0),5) για υπερχείλιση
⋮		⋮ το τραβάω
1000		

→ Αστοχία κάλυψης αναγκών = Countif (όλα τα S χωρίς S_0, "=0") / count (S) → 1000

→ Αξιοπιστία = 1 - αστοχία

→ Πιθανότητα υπερπήδησης = Countif (όλα τα S , "=5") / Count (S)

→ Αν κάνω την ζήτηση 9 η αστοχία πρέπει να αυξηθεί και η υπερπήδηση επίσης. Αν μειώσω την ζήτηση πρέπει να δω το ανάποδο

γ) Για την καμπύλη πρέπει να καλύψουμε τα (5,10) της ζήτησης κατά 0,1

βάζουμε $S_{5,1}$ $S_{5,2}$ $S_{5,3}$ ...

και τραβάω δεξιά => \$ στο μπροστά (σε εισροή και ζήτηση)

→ και αντίστοιχα τραβάμε αστοχία, αξιοπιστία, υπερπήδηση, ζήτηση

→ κάνουμε διάγραμμα x = ζήτηση (από 5 έως 10 οι άξονες)

y = αξιοπιστία (από 0 έως 1 οι άξονες)

και βρίσκουμε το 5%

δ) το τρέχουμε ξανά για 5+30%·5

Κουτσογιάννης:

→ η δομική αστοχία είναι πολύ πιο σημαντική
(η) λειτουργική αστοχία → 2 κατηγορίες

→ οι δεκτές πιθανότητες αστοχίας είναι 1-10% πχ σε αρδευτικά έργα.
↳ σε ύδρευση δεν μπορεί να γίνει αυτό μόνο 1% αστοχία

→ όταν πλησιάζουμε στην αστοχία λαμβάνουμε μέτρα για να μην συμβεί.

→ τρελή απλοποίηση ότι η εισροή = σταθερή. και πήραμε κανονική κατανομή στην εισροή
↓ μόνο σε ετήσια θα μπορούμε
↓ πήραμε ανεξαρτησία τιμών = αδύνατο. => πρέπει να το μελετήσουμε.

Παράδειγμα καθηγητή

→ Κάποιος πάει ταξίδι στην Σκωτία για 2 μέρες. Δεν υπάρχει εξάρτηση βροχής.

α) Ποιά η πιθανότητα να έβρεχε και τις 2 μέρες, αν στην 1 μέρα είναι 50%
Αυτό ισχύει και στα επόμενα

$P(\alpha) = \frac{1}{2} \cdot \frac{1}{2} = 0,25 = 25\%$

β) Βάζει μια φωτογραφία και κρατάει ομπρέλα και είναι η 1η μέρα (άρα βρέχει) ποια η πιθανότητα να έβρεξε και τις 2 μέρες.

$P(\beta) = 1 \cdot \frac{1}{2} = 0,5 = 50\%$

γ) Έβαλε την φωτό αφού γύρισε για 1 φωτογραφία με ομπρέλα την 1 εκ των 2 ημερών ποιά η πιθανότητα να έβρεξε και τις 2 μέρες.
↓ Δεν ξέρω ποιά μέρα

1η	2η		Για τυχαία μεταβλητή X	
Β	Β	← βρέχει και τις 2 μέρες αλλά έχω 3 δυνατά σενάρια βροχής ⇒ $P(\gamma) = 1/3$	1	1
Β	Δ		1	0
Δ	Β		0	1
Δ	Δ	→ Απορρίπτω σίγουρα αυτό (γιατί δεν βρέχει)	0	0

→ στην επιστήμη τα περισσότερα είναι υποκειμενικά και σχετίζονται με τις πληροφορίες (πχ άλλο σύστημα παρατηρητή)

Ανεξάρτητα γεγονότα: $P(AB) = P(A) \cdot P(B)$

Εξαρτημένα γεγονότα:

$P(A|B) := \frac{P(AB)}{P(B)}$ => Θεώρημα Bayes: $P(B|A) = P(B)\frac{P(A|B)}{P(A)}$

P(A) δεδομένου B — εξορισμού — ανάλυση

Δεσμευμένη πιθανότητα

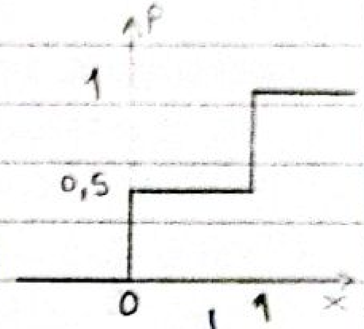
→ κατεβάζουμε το βιβλίο ή το παίρνουμε απ' την όλγα. Διαβάζουμε το κεφάλαιο 2 και 3.
↓ αυτά θα προσπαθήσει να μας πει.

→ η διασπορά είναι η ροπή αδράνειας.

Συνάρτηση κατανομής: $F(x) := P(\underline{x} \le x)$

↓ τυχαία μεταβλητή (0 ή 1) για το προηγούμενο παράδειγμα

↳ νούμερο τυχαίο $(0,+\infty)$



→ η συνάρτηση κατανομής βγαίνει μια σκαλίτσα. (στο 0 και στο 1)

↓ διακριτή μεταβλητή

→ μας νοιάζει και πόσο έβρεξε.

Πυκνότητα πιθανότητα: $f(x) := \frac{dF(x)}{dx}$ και $f(x) = \lim_{\Delta x \to 0} \frac{P(x \le \underline{x} \le x+\Delta x)}{\Delta x}$

→ Διακριτή μεταβλητή = σκαλιά

→ Συνεχή μεταβλητή = τσουλήθρα

→ και σκαλιά και τσουλήθρα = μεικτή μεταβλητή.

→ Από το διάγραμμα της συνάρτησης κατανομής φαίνεται και η πυκνότητα πιθανότητας

↓ ουσιαστικά η κλίση της τσουλήθρας (sos)

ή γράφω $\frac{\Delta F}{\Delta x}$

πιο κοντά στο 0 είναι πιο πιθανό γιατί είναι πιο πυκνή η πυκνότητα πιθανότητας

→ Θυμόμαστε τις ροπές απ' τις πιθανότητες

Μέση τιμή = μ , Διασπορά = σ^2

Πιθανότερη τιμή = x_p

Διάμεση τιμή = $x_{0,5}$

Τυπική απόκλιση = σ

Συντελεστής διασποράς = $C_v = \sigma/\mu$

Συντελεστής ασυμμετρίας = $C_s = \mu^{(3)}/\sigma^3$

Εντροπία: Δηλώνει αβεβαιότητα ή αταξία; Είναι ισοδύναμα;

→ η εντροπία πάντα αυξάνεται. => 2ος θερμοδυναμικός νόμος, δεν διατηρείται, υπάρχει αλλαγή και έχουμε ζωή.

→ αναμενόμενη τιμή

⇓

το άθροισμα.

Αναμενόμενες τιμές:

→ γενικά την λέμε μέση τιμή όταν μιλάμε για συγκεκριμένο x, αλλά επειδή μπορεί να μιλάμε και για μια συνάρτηση (πχ x^2) χρησιμοποιούμε τον όρο αναμενόμενη τιμή της συνάρτησης αυτής.

$E[g(x)] = \sum_{j=1}^{n} g(x_j)\cdot P(x_j)$, για διακριτή μεταβλητή

το άθροισμα του γινομένου των πιθανοτήτων με την τιμή του x.

$E[g(x)] = \int_{-\infty}^{+\infty} g(x)\cdot f(x)\,dx$, για συνεχή μεταβλητή

→ Αν $g(x)=x$ ⇒ Μέση τιμή
Αν $g(x)=(x-\mu)^2$ ⇒ Διασπορά
Αν $g(x)=(x-\mu)^3$ ⇒ 3η ροπή κτλ.

προφανώς όλα $\cdot p(x_j)$ και Σ

συμβολίζεται $\mu^{(3)}$

Εντροπία (συνέχεια):

→ μια επιπλέον ροπή είναι η εντροπία.

→ όλοι οι φυσικοί νόμοι είναι διατήρησης, πχ ενέργειας, ορμής, μάζα, φορτίο

όταν διατηρείται κάτι δεν έχω αλλαγή ⇒ καήκαμε
γιατί ο άνθρωπος είναι προϊόν αλλαγής.

→ η δύναμη που ωθεί αυτή την αλλαγή είναι η εντροπία, ο 2ος νόμος. **Η εντροπία ενός συστήματος αυξάνεται.**

→ η αταξία έχει αρνητική έννοια, ενώ η αβεβαιότητα όχι τόσο (η ουσία ενός αγώνα βόλεϊ είναι ότι δεν ξέρεις ποιός θα νικήσει)

όσο πιο απρόβλεπτο τόσο πιο συναρπαστικό.

→ τον ταμιευτήρα τον κάνουμε για να μειώσουμε την αβεβαιότητα, όχι να την εξαλείψουμε.

→ όταν μπορώ να σε προβλέψω δεν έχεις επιλογή ⇒ δεν έχεις και ελευθερία.

→ η εντροπία και η ελευθερία είναι απόλυτα συμβατές

→ όταν υπάρχουν εξαρτήσεις δίνεται νόημα στην ελευθερία. όχι σκέτο rand()

Ορισμός: Και η εντροπία είναι αβεβαιότητα: $\Phi[x] = E[-\ln P(x)] = \sum_{j=1}^{n} -\ln P(x_j)\cdot P(x_j)$

ποσοτικοποίηση της αβεβαιότητας είναι ουσιαστικά

αναμενόμενη τιμή του λογαρίθμου της πιθανότητας.

→ άρα η μεγιστοποίηση της εντροπίας μας λέει ότι η αβεβαιότητα συνεχώς αυξάνεται (σε ένα μεγάλο σύστημα). Σε ένα τμήμα μπορεί να μειώνεται.

→ ο Ήλιος αυξάνει την εντροπία του συνεχώς, για να μπορεί η Γη να την κρατάει σταθερή
→ αν πάρουμε την εντροπία σε ένα ζάρι (που έχει σε όλες τις καταστάσεις ίδια πιθανό-τητα το $\frac{1}{6}$):

$$\varphi = \sum_1^6 \left(-\ln\frac{1}{6}\right)\left(\frac{1}{6}\right) = \sum_1^6 \ln 6 \cdot \frac{1}{6} = \frac{\cancel{6}}{\cancel{6}} \ln 6 = \ln 6$$

πιθανές καταστάσεις

ο Shannon 2)
→ μετά βρήκε ότι ο τύπος που περιγράφει την πληροφορία είναι ίδιος με την εντροπία.

1) ο Boltzmann είπε ότι εντροπία είναι ίση με τον λογάριθμο των πιθανών καταστάσεων

(*) Έχω την εντροπία που είναι μέτρο της αβεβαιότητας, μόλις αυτό γίνει βεβαιότητα η εντροπία γίνεται μηδέν. Άρα αυτό που είχα ως αβεβαιότητα είναι ταυτόχρονα και το μέτρο που δείχνει την πληροφορία, γιατί με την πληροφορία μηδένισα την εντροπία.

3) ο Jaynes:
→ Εισάγει την αρχή της μέγιστης εντροπίας. Δηλαδή σε ένα σύστημα αφού λάβουμε υπόψιν ότι ορίζεται από ανθρώπινους νόμους και το γράψουμε σε μορφή περιορισμών, ότι απομένει (το υπόλοιπο) είναι τέτοιο ώστε να μεγιστοποιείται η εντροπία.

οι περιορισμοί όμως παίζουν έναν ρόλο.

→ distribution = κατανομή

	Name	Probability density function	Distribution function
1)	Uniform in [0, 1]	$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$	$F(x) = \max(0, \min(x, 1))$
2)	Exponential	$f(x) = \begin{cases} e^{-x/\mu}/\mu & \text{for } x \ge 0 \\ 0 & \text{for } x < 0 \end{cases}$	$F(x) = \begin{cases} 1 - e^{-x/\mu} & \text{for } x \ge 0 \\ 0 & \text{for } x < 0 \end{cases}$
3)	Normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du$

1) Η τυχαία μεταβλητή $\underline{x}$ κυμαίνεται στο 0 και 1, τίποτα άλλο, χωρίς κανέναν άλλο περιορισμό τότε η αρχή της μέγιστης εντροπίας θα μας οδηγήσει στο rand(), δηλαδή μια ομοιόμορφη κατανομή στο [0, 1]

3) Αν μιλάμε για τα μόρια του αέρα έχουμε την αρχή διατήρησης της ενέργειας $= m \cdot u^2$ η $E[u^2]$ πρέπει να κρατηθεί, διατηρηθεί. ⇒ το βάζουμε σαν περιορισμό και βγάζουμε την $f(x)$ από πάνω που είναι η κανονική κατανομή

2) Αν $E[u]$, του ίδιου του μεγέθους, θέλουμε να μείνει σταθερό προκύπτει η εκθετική κατανομή.

Κεντρικό οριακό θεώρημα:

→ Όταν προσθέτεις πολλά, ανεξάρτητα από το τι θα είναι αυτά, η κατανομή τείνει στην κανονική κατανομή που είδαμε πριν.

→ βροχή σε ημερήσια κλίμακα μοιάζει με εκθετική κατανομή (μοιάζει με Παρέτο → πιο χοντρή η ουρά της)

↓

ορίζουμε κλίμακα για να πάμε τι κατανομή ακολουθεί.

(ξεκινάμε απ' την στιγμιαία κλίμακα και πάμε σε πιο μεγάλες)

→ οι περιορισμοί λαμβάνουν υπόψη την πληροφορία που έχουμε εμείς.

4-11-2021

Στοχαστικές μέθοδοι
Κουτσογιάννης
Μάθημα 4ο

→ υπάρχουν 4 έννοιες της διασποράς
Table 4.1 σελ 145 Βιβλίο Κουτσογιάννη

1) Θεωρητική ή αληθής διασπορά:

→ πρέπει να ξέρουμε την pdf

Έχουμε ένα ζάρι ⇒ αντιστοιχεί σε μια ομάδα με 6 κελιά. (διακριτές τιμές)

θα ήταν ένα περίεργο ζάρι που δεν θα εκφράζει κάτι απόμακρο ⇒ πρέπει να ακολουθεί την αρχή της μέγιστης εντροπίας.

x	P (πιθανότητα)	= τιμή κελιού / sum κελιών	= −ln (κελιού)	= $P_i \cdot (-\ln P_i)$ = εντροπία
1	πχ 0,5			
2	0,7	= 0,7 / 2,6	ουσιαστικά −ln(P)	0,353
3	0,1	= 0,1 / 2,6	3,258	0,125
4	0,1	⋮	3,258	0,125
5	0,9	0,346	1,061	0,367
6	0,3	0,115	2,159	0,249
sum =	2,6 δεν μου κάνει	1		1,537

πρέπει να κάνει 1 — μου κάνει

Εντροπία: $\Phi[\underline{x}] = -\sum P_i \cdot \ln P_i$

Πάμε στο solver και το θέλουμε max με αλλαγή της 2ης στήλης.

Αυτό το κατάλαβα, πρέπει να μου το κάνει μέγιστο σύμφωνα με την αρχή της μέγιστης εντροπίας
sum = 1,792

γιατί έτσι μέγιστοποιείται η εντροπία ⇐ βγάζει $P = 1/6$ της κάθε μίας (χωρίς να του το πούμε εμείς)

→ Για να βρω μέση τιμή $= \sum x \cdot P(x)$

→ Για να βρω την διασπορά: $\boxed{\gamma = (x-\mu)^2 \cdot P(x)}$ (για διακριτή μεταβλητή)

x	P(x)	μ	γ
1	0,167	1·0,167	$(1-3,5)^2 \cdot 0,167$
2	0,167	2·0,167	$(2-3,5)^2 \cdot 0,167$
3	0,167	⋮	⋮
4	0,167		
5	0,167		
6	0,167	6·0,167	$(6-3,5)^2 \cdot 0,167$
sum		$\boxed{3,5}$ μέση τιμή	$\boxed{2,92}$ διασπορά

Θεωρητική ή αληθής διασπορά για να την βρω πρέπει να ξέρω τις πιθανότητες.

2) Εκτίμηση της διασποράς από τα δεδομένα

(Εκτίμηση = αριθμός)

→ τώρα έχουμε δεδομένα.

→ δουλεύουμε πάλι με την περίπτωση του ζαριού

Πείραμα	Αποτέλεσμα (τυχαίο)
1	=round(rand()·6+0,5 , 0)
2	
3	
⋮	
100	

στρογγυλεύση τιμές από 0-1 / για να μην βγάλει ψηφία για μηδέν στρογγύλευση / γιατί έχουμε 6 σωστά αποτελέσματα / → Δεκαδικά

→ τώρα έστω ότι θέλω να βρω την μέση τιμή ⇒ average(2η στήλη) = 3,53 ≃ όσο και πριν

→ να δούμε και την διασπορά ⇒ var(2η στήλη) = 2,77 ≃ 2,92 (κοντά πάλι) και δεν αλλάζει πολύ

→ όσο πιο μεγάλο δείγμα διάλεγα τόσο πιο κοντά θα ήμουν στις προηγούμενες τιμές

3) Εκτιμήτρια της διασποράς

(Εκτιμήτρια = τυχαία μεταβλητή)

→ μοιάζει αρκετά με την 2η διασπορά (έχει − από κάτω ⇒ Δεν θεωρούνται νούμερα, αλλά τυχαίες μεταβλητές)

4)

→ ο αριθμός των επαναλήψεων να γίνει άπειρος (είναι μια θεωρητική έννοια)

→ αν αυτό το lim( ) γίνει ίσο με την πραγματική, τότε λέμε ότι η τυχαία μας μεταβλητή χαρακτηρίζεται από εργοδικότητα

→ στις εφαρμογές μας θα υπάρχει αυτή η υπόθεση της εργοδικότητας.

→ αν έχω πολλές ζαριές μπορώ να γράψω $\underline{X}_1$ (1η ζαριά), $\underline{X}_2$ (2η ζαριά), $\underline{X}_3$ (3η ζαριά) κτλ

α) η μ.τ. της $\underline{X}_1$ και $\underline{X}_2$ είναι ίδια ή διαφορετική; ίδια

$\underline{X}_1$ =3,5, $\underline{X}_2$ =3,5 γιατί εκεί μεγιστοποιείται η εντροπία

↓ θα έχουν ίδια μέση τιμή και ίδια διασπορά.

β) Εξαρτάται η έκβαση της 2ης ζαριάς απ' την 1η; όχι (το ξέρουμε διαισθητικά)

↳ θα έχουμε 36 πιθανότητες και το τρέχουμε με τον ίδιο τρόπο με πριν

→ τώρα σταματάμε να τα λέμε ζαριές και $\underline{X}_1$ = παροχή ποταμού Αχελώου Γενάρη 2001

$\underline{X}_2$ = —"— Φλεβάρη 2001

α) το $\underline{x_1}$ με το $\underline{x_7}$ έχουν ίδια μέση τιμή; όχι, δεν έχουν ίδια μέση τιμή
Γενάρης Ιούλιος

β) το $\underline{x_1}$ με το $\underline{x_2}$ είναι ανεξάρτητα; όχι, εξαρτώνται

Συμπέρασμα: Όταν μιλάω για ένα φυσικό φαινόμενο (όπως η παροχή του ποταμού) πρέπει να χρησιμοποιήσω διαφορετικές τυχαίες μεταβλητές → Στοχαστική Ανέλιξη
(stochastic Process)
Διεργασία (φυσικό) Ανέλιξη (μαθηματικό)

→ η ανέλιξη μεταβάλλεται στο χρόνο είναι $\underline{x}(t)$, μπορούμε να διακριτοποιούμε την μεταβλητή χρησιμοποιώντας μια ισοδιάσταση (μέση τιμή χρονικά στη διάρκεια ενός μήνα)
πχ 1 μήνας

→ η ανέλιξη δεν έχει γεωμετρική απεικόνιση, γιατί είναι μια μαθηματική έννοια. Αυτό που σχεδιάζω είναι την υλοποίηση μιας ανέλιξης
δειγματοσυνάρτηση ή χρονοσειρά
(αν είναι σε διακριτό χρόνο)

→ το ζάρι είναι μια διακριτή μεταβλητή και στον χρόνο, αλλά και στην κατάσταση
1 ή 2 ή 3 ζαριές — θα δώσει 1 ή 2 ή ... ή 6

→ η παροχή (του ποταμού) είναι μια συνεχής μεταβλητή και στον χρόνο, αλλά και στην κατάσταση
χρόνος 1, 1.5, 2 — υπάρχει η τιμή 10, 10.1 κτλ

Άρα 2 μορφές διακριτότητας και συνέχειας:
στον χρόνο — στην κατάσταση

→ αν έρθω στην περίπτωση της ζαριάς μπορώ να έχω κάτι ανάλογο σε συνεχή χρόνο; ΝΑΙ
Λευκός θόρυβος (είναι μια ανέλιξη)
κεφάλαιο 3.10 (βιβλίου)
↳ σε συνεχή χρόνο που μοιάζει με ζαριά.

→ τελείως ιδεατό μαθηματικό μοντέλο, δεν μπορεί να την σχεδιάσει κανείς.

→ έχει μια μέση τιμή σταθερή συνεχώς, συσχέτιση 0 στον χρόνο t

Δηλαδή αν έχω 2 μεταβλητές ορίζω το μέγεθος που λέγεται covariance = COV

άρα $\mathrm{cov}(\underline{x_1}, \underline{x_2}) =: E[(\underline{x_1}-\mu_1)(\underline{x_2}-\mu_2)] =$

$$= \int_{-\infty}^{+\infty} (x_1-\mu)(x_2-\mu_2) \cdot f(x_1,x_2)\,dx_1\,dx_2$$

σαν αποτέλεσμα θα έχει νούμερο.

→ Λευκός θόρυβος θα πει:

$$cov\left[\underline{x}(t),\ \underline{x}(t+dt)\right]=0 \quad \text{(ασυσχέτιστα)}$$

→ Η συνδιασπορά ενός πράγματος με τον εαυτό του λέγεται διασπορά, άρα:

$$cov\left[\underline{x}(t),\ \underline{x}(t)\right]=\infty$$

→ κάθε στοχαστική ανέλιξη έχει και ένα φάσμα, το φάσμα ισχύος.
→ ισχύς συναρτήσει συχνότητας.

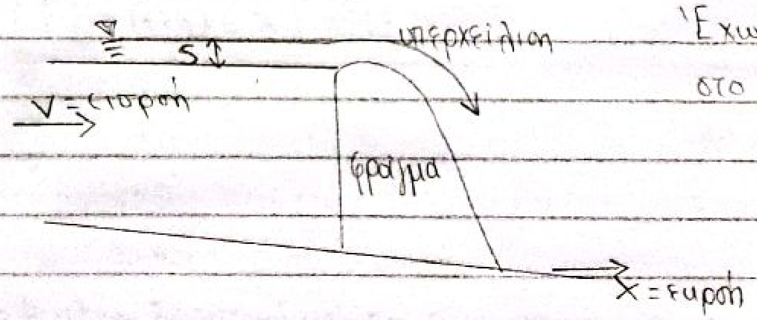
"το φάσμα του λευκού θορύβου είναι μια ευθεία γραμμή."

→ στο έτος εξαλείφονται όλες οι περιοδικότητες.

Στάσιμη ανέλιξη: Όταν τα στατιστικά χαρακτηριστικά δεν μεταβάλλονται στον χρόνο (υπάρχει όμως μεταβολή της φυσικής διεργασίας στο χρόνο)

→ Άλλο μόνιμη = σταθερά στο χρόνο ≠ στάσιμη = όλα μεταβαλλόμενα στο χρόνο αλλά με σταθερά στατιστικά χαρακτηριστικά
↓
μέση τιμή, διασπορά, συσχέτιση

Παράδειγμα καθηγητή



Έχω και αποθήκευση (s) που μεταβάλλεται στο χρόνο ⇒ $\boxed{V-X=\dfrac{ds}{dt}}$ (1)

λόγω υδραυλικής: $X=g(s)$
και στην απλούστερη περίπτωση, κάνοντας μια γραμμικοποίηση $X=k\cdot S$ (2)

Από (1) και (2) προκύπτει: $v(t)-x(t)=\dfrac{1}{k}\dfrac{dx}{dt}$, έστω $a=\dfrac{1}{k}$

Γραμμική διαφορική εξίσωση 1ης τάξης με σταθερούς συντελεστές

Διαφορική εξίσωση ⇒ $\boxed{a\dfrac{dx}{dt}+x(t)=v(t)}$

(με οριακή συνθήκη $x(0)=x_0$)

υποθέτω ότι είναι γνωστό ότι μπαίνει στον ταμιευτήρα μου.

→ για να λύσουμε την διαφορική εξίσωση πάμε στο wolframalpha

$$x(t)=e^{-t/a}\left(\int_0^t \frac{e^{\zeta/a}v(\zeta)}{a}\,d\zeta+x_0\right)$$

⇒ Αυτό μας λέγε τα υλικά και μαθηματικά.

→ μπορούμε να τις κάνουμε και τυχαίες μεταβλητές.

→ Για τον χρόνο h έχουμε: $\underline{x}(t+h)=\underline{x}(t)\,e^{-h/a}+\dfrac{1}{a}\displaystyle\int_t^{t+h}\underline{v}(\zeta)\cdot e^{(\zeta-t-h)/a}\cdot d\zeta$

→ Αυτό που μας λέει η λύση της δ.ε. είναι ότι: η μελλοντική τιμή $x(t+h)$ εξαρτάται από το $v(t)$ παρόν, αλλά όσο απομακρυνόμαστε στο μέλλον τόσο μικρότερη γίνεται η εξάρτηση αυτή $e^{-h/k}$.
Άρα αν το h γίνει πολύ μεγάλο η εξάρτηση από το παρόν μηδενίζεται.
↓ Αυτό είναι ως προς το ίδιο το φαινόμενο

→ Εξαρτάται και απ' την εισροή $v(t)$ (στο διάστημα $t \to t+h$)

→ η βροχή είναι τυχαία, σχεδόν σαν λευκός θόρυβος ⇒ παραδοχή (χονδροειδής)
$v(t)$ = λευκός θόρυβος

→ το $x(t)$ με το $v(t)$ είναι στοχαστικά ανεξάρτητα μεταξύ τους.
εισροή σε προηγούμενο χρόνο
σε χρόνους μετά το παρόν $t \to t+h$ (εισροή σε επόμενο χρόνο)

→ δεν μας νοιάζει τι έγινε σε προηγούμενο χρόνο

Ανέλιξη Μάρκοφ:

→ ότι κάναμε με το μοντέλο του υπερχειλιστή είναι μια μορφή αυτής.

→ Ρώσος μαθηματικός, αρχές 20ου αιώνα

$$\boxed{P[\underline{x}(t+h) \le x \mid x(t), x(t_n), \ldots, x(t_1)] = P[\underline{x}(t+h) \le x \mid x(t)]}$$

⇓ Ερμηνεία:
η πιθανότητα να συμβεί κάτι στον χρόνο $t+h$, δεδομένου ότι γνωρίζω τι έχει συμβεί στο παρελθόν και στο παρόν, είναι ίση με την πιθανότητα να συμβεί κάτι σε χρόνο $t+h$, δεδομένου ότι γνωρίζω μόνο το παρόν.

→ Η πλέον δραστική υπόθεση είναι να πω ότι το μέλλον δεν εξαρτάται καθόλου ούτε από το παρόν ούτε από το παρελθόν ⇒ Λευκός θόρυβος.

→ η απόλυτη ανεξαρτησία δεν ισχύει για τα ποτάμια, τους σεισμούς

→ μπορούμε να κάνουμε και άλλη απλοποιητική παραδοχή.

$\boxed{\underline{x}_i = a \cdot \underline{x}_{i-1} + \underline{v}_i}$ Μοντέλο αυτοπαλινδρόμησης AR(1) (προκύπτει από το autoregression)
(σε διακριτό χρόνο Μάρκοφ)

→ τα ακραία υποχωρούν προς το μέσον ⇒ βιολογικός νόμος
υποχώρηση = regression ← Λάθος θεώρηση

→ ισχύει ο γενικός κανόνας: $E[a\underline{x}+b\underline{y}] = aE[\underline{x}] + bE[\underline{y}]$

Άρα $E(\underline{x}_t) = E[a\underline{x}_{t-1} + \underline{v}_t] = aE[\underline{x}_{t-1}] + E[\underline{v}_t]$, αφού έχω στασιμότητα $E(x_t) = E(x_{t-1})$

$m = a \cdot m + m_v \Leftrightarrow \boxed{m_v = (1-a) \cdot m}$ Βρίσκω το m_v

μέση τιμή

$a < 1 \Rightarrow a = \rho$ = συντελεστής συσχέτισης

→ Έχω τώρα τις σχέσεις: $\underline{x}_t = \alpha \cdot \underline{x}_{t-1} + v_t$ και $m = \alpha \cdot m + m_v$ και τις αφαιρώ κατά μέλη =>

(1) $\underline{x}_t - m = \alpha(\underline{x}_{t-1} - m) + \underline{v}_t - m_v \Leftrightarrow$

$(\underline{x}_t - m)(\underline{x}_{t-1} - m) = \alpha \cdot (\underline{x}_{t-1} - m)^2 + (\underline{v}_t - m_v)(\underline{x}_{t-1} - m) \Leftrightarrow$

επειδή είναι ανεξάρτητα $E[v_t - m] E[x_{t-1} - m]$

$E[(\underline{x}_t - m)(\underline{x}_{t-1} - m)] = \alpha E[(\underline{x}_{t-1} - m)^2] + E[(\underline{v}_t - m_v)(\underline{x}_{t-1} - m)]$

ή x_{t-1} (πάσιμο)

με αυτόν τον τρόπο μπορώ να υπολογίσω την παράμετρο α.

$\boxed{cov[\underline{x}_t, \underline{x}_{t-1}] = \alpha \cdot var(\underline{x}_t) + \emptyset}$

συνδιασπορά διασπορά

μπορεί να διαγραφεί αυτός ο όρος.

(*) $E[\underline{v}_t - m] = E[\underline{v}_t] - m_v = m_v - m_v = \emptyset$

→ Παίρνω την (1) και την υψώνω στο τετράγωνο.

$(\underline{x}_t - m)^2 = \alpha^2 \cdot (\underline{x}_{t-1} - m)^2 + (\underline{v}_t - m_v)^2 + 2 \cdot \alpha \cdot (\underline{x}_{t-1} - m) \cdot (\underline{v}_t - m_v) \Leftrightarrow$

$E[(\underline{x}_t - m)^2] = \alpha^2 \cdot E[(x_{t-1} - m)^2] + E[(\underline{v}_t - m_v)]^2 + 2\alpha \cdot E[(x_{t-1} - m)(\underline{v}_t - m_v)]$

το συμβολίζουμε με σ^2 ή σ^2 = η διασπορά

x_{t-1} (στασιμότητα)

$\sigma^2(x) = \alpha^2 \sigma^2(x) + \sigma_v^2(x) + \emptyset$

$\boxed{\sigma_v^2 = (1 - \alpha^2)\sigma^2}$

→ τα 3 μπλε χρειάζομαι για την προσομοίωσή μου.

<u>Παράδειγμα καθηγητή</u>

Έχουμε $m = 10$

$\sigma^2 = 20$

$p = 0{,}5$ (συντελεστής αυτοσυσχέτισης) με $p = \frac{\sigma xy}{\sigma x \cdot \sigma y}$, με $\sigma xy = cov$

με $-1 < p < 1$ (sos)

<u>Λύση:</u>

$\sigma = \sqrt{20}$ => δε εμάς x και y τώρα είναι ίδιο, άρα:

$\sigma xy = p \cdot \sigma x \cdot \sigma y = 0{,}5 \cdot \sqrt{20} \cdot \sqrt{20} = 10 = cov$

- για να βρω το α χρησιμοποιώ την σχέση: $\sigma xy = \alpha \cdot var(x)$

  άρα $\alpha = \sigma xy / var = 10 / \sigma^2 = 10/20 = 0{,}5$ (έπρεπε να μας κόψει ότι $\alpha = p$)

- για να βρω το m_v χρησιμοποιώ την σχέση: $m_v = (1 - \alpha) \cdot m$

  άρα $m_v = (1 - 0{,}5) \cdot 10 = 5$

- για να βρω το σ_v^2 χρησιμοποιώ την σχέση: $\sigma_v^2 = (1-\alpha^2)\cdot\sigma^2$

  άρα $\sigma_v^2 = (1-0{,}5^2)\cdot 20 = 15$

→ Έστω ότι έχω κανονική κατανομή και 10.000 νούμερα, τους τυχαίους αριθμούς τους βγάζω με norminv(rand(), $\underset{m_v}{5}$, $\underset{\sigma_v}{15^{0,5}}$)

Προσομοίωση	Τυχαίοι αριθμοί V	$x_0 = 10$ και εφαρμόζω το μοντέλο
1	=norminv(rand(),5,$15^{0,5}$)	$= \underset{0,5}{\alpha}\cdot X_0 + \varepsilon^{\text{αριστερά}}$ $(x_t = \alpha x_{t-1} + v_t)$
2	⋮	
3		
⋮		
10.000		

=average(2ης στήλης) = 5,05 ✓ => 3ης = 9,96 ✓

= var(2ης στήλης) = 15,03 ✓ => 3ης = 19,46 ✓

παρατηρώ ότι το μοντέλο λειτουργεί πάρα πολύ καλά.

→ για να δω ότι το μοντέλο δουλεύει σωστά πρέπει να βγάλω συντελεστή αυτοσυσχέτισης 0,5

= correl(3ης στήλης, 3ης στήλης κατά 1 προς τα κάτω) = 0,509

→ η μεγάλη συσχέτιση υποδηλώνει την ύπαρξη μοτίβου.

→ άλλα 2 μοντέλα που μπορούμε να δούμε είναι το ARMA(1,1) και AR(2)

→ η βασική ύλη είναι κεφάλαιο 2,3,4.

11-11-2021

Στοχαστικές Μέθοδοι
Ανυυ , Δημητριάδης
Μάθημα 5ο

Στάσιμα στοχαστικά μοντέλα:

Άσκηση 2α:

→ συνδιασπορά 2 αριθμών $\underline{x}, \underline{y} = cov[\underline{x},\underline{y}] = E[(\underline{x}-m_x)(\underline{y}-m_y)] = E[\underline{x}\underline{y}] - E[\underline{x}] - E[\underline{y}]$

→ αυτοσυσχέτιση $= \rho_{xy} = \dfrac{cov[x,y]}{\sqrt{Var[x]\cdot Var(y)}} = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$ με $\boxed{-1 \le \rho_{xy} \le 1}$ (sos)
↳ τυπική απόκλιση

→ διασπορά $= \sigma^2 = E[(\underline{x}-m)^2] = var(x)$

→ μέση τιμή $= m = E(\underline{x})$

→ όσο μεγαλώνει η τάξη της ροπής, τόσο πιο δύσκολη γίνεται η εκτίμηση

→ pdf: "Στάσιμα στοχαστικά μοντέλα μιας μεταβλητής"
"Τυχαίες μεταβλητές, στοχαστικές ανελίξεις και χρονοσειρές"
AR(1) σελ 7 ARMA(1,1) σελ 9

Μοντέλο AR(1):

$\underline{x_i} = a \cdot \underline{x_{i-1}} + \underline{v_i}$ (1) → τρόπος γένεσης της χρονοσειράς

→ το πιο απλό είναι να βρεις την μέση τιμή ⇒ περνάς απλώς αναμενόμενες τιμές.

$E[\underline{x_i}] = a \cdot E[\underline{x_{i-1}}] + E[\underline{v_i}] \Leftrightarrow$ ισχύει η ιδιότητα της γραμμικότητας, γι'αυτό μπορώ να το σπάσω

$\mu_x = a \cdot \mu_x + \mu_v \Leftrightarrow$ λευκός θόρυβος

$\mu_v = (1-a)\mu_x$ ή με $E[\underline{x_i}] = E[\underline{x_{i-1}}]$ λόγω στασιμότητας

$\boxed{\mu_x = \dfrac{\mu_v}{1-a}}$ Μέση Τιμή (δεν επηρεάζεται από τον χρόνο για οποιαδήποτε χρονική μετατόπιση)
↓
υστέρηση

→ Αυτός είναι ένας στοχαστικός αλγόριθμος, που μέσα απ' το μοντέλο που κάνει παλινδρόμηση στην προηγούμενη τιμή του x (προσθέτοντας έναν όρο λευκού θορύβου) μπορείς να φτιάξεις μια ανέλιξη AR(1) ⇒ Ανέλιξη που διατηρεί μόνο την αυτοσυσχέτιση 1ης τάξης

↳ Πειράζοντας την μ_v αλλάζει και η μ_x

→ Αφαιρώ τις 2 εξισώσεις κατά μέλη:

$\underline{x_i} - \mu_x = a(\underline{x_{i-1}} - \mu_x) + \underline{v_i} - \mu_v \Leftrightarrow$

Θέτω $\underline{x_i} - \mu_x = x_i'$ ⇒

$x_i' = a \cdot x_{i-1}' + v_i' \Leftrightarrow$

$x_i'^2 = a^2 \cdot x_{i-1}'^2 + v_i'^2 + 2 \cdot a \cdot \underbrace{x_{i-1}' \cdot v_i'}$ ⇔ βάζω αναμενόμενες τιμές

Ανεξάρτητες, γιατί από την (1) το x_{i-1} θα εξαρτιόταν από το v_{i-1} και όχι από το v_i (κοιτάμε τους δείκτες)

Διασπορά του Λευκού θορύβου ↙ Ανεξάρτητα (αιτιολογία)

$E[x_i'^2] = a^2 \cdot E[x_{i-1}'^2] + E[v_i'^2] + 2a\, \underbrace{E[x_{i-1}']}_{} \cdot E[v_i] \Leftrightarrow$

$\sigma_x^2 = a^2 \cdot \sigma_x^2 + \sigma_v^2 + \varnothing \Leftrightarrow$

Λόγω στασιμότητας

$E[(x_{i-1} - \mu_x)] = E(x_{i-1}) - \mu_x = \mu_x - \mu_x = 0$

$\sigma_v^2 = (1-a^2)\sigma_x^2 \Leftrightarrow$ Συμβολίζουμε $\sigma_x^2 = C_0$

$\boxed{\sigma_v^2 = C_0 \cdot (1-a^2)}$ Διασπορά

Διασπορά x, αυτοσυνδιασπορά (την συνδιασπορά της x με τον εαυτό της) για υστέρηση Ø

$= \dfrac{\text{αυτοσυνδιασπορά}}{\text{διασπορά}}$

→ Τώρα θέλω να υπολογίσω την αυτοσυσχέτιση:

(προσπαθώ με προσθαφαιρέσεις να δημιουργήσω τον ορισμό για το μέγεθος που θέλω)

$\underline{x_i} = a \cdot \underline{x_{i-1}} + \underline{v_i} \Leftrightarrow$

$\underline{x_i'} = a\, \underline{x_{i-1}'} + \underline{v_i'} \Leftrightarrow$ πολλαπλασιάζω με $\underline{x_{i-1}'}$

$\underbrace{\underline{x_i'} \cdot \underline{x_{i-1}'}}_{} = a \cdot \underline{x_{i-1}'^2} + \underline{v_i'} \cdot \underline{x_{i-1}'} \Leftrightarrow$

↳ Ανεξάρτητα

$(x_i - \mu_x)(x_{i-1} - \mu_x)$

↳ με υστέρηση $\tau = 1$

$E[\underline{x_i'} \cdot \underline{x_{i-1}'}] = a \cdot E[\underline{x_{i-1}'^2}] + E[\underline{v_i'}] \cdot E[\underline{x_{i-1}'}] \Leftrightarrow$

$C_1 = a \cdot C_0 + \varnothing \Leftrightarrow$

Αυτοσυνδιασπορά για υστέρηση 1

$\boxed{C_1 = a \cdot C_0}$ **Αυτοσυνδιασπορά για υστέρηση 1 της x**

↕ Εκφράζουν τις αλλαγές του μοντέλου στο χρόνο που είναι συσχετισμένες

$\rho_1 = \dfrac{C_1}{C_0} \Leftrightarrow \boxed{\rho_1 = \dfrac{C_1}{C_0} = a}$ **Αυτοσυσχέτιση για υστέρηση 1**

Από τον ορισμό Αυτοσυσχέτιση

Άρα στο AR(1) => $\rho_1 = a$

<u>HW</u> Βγάζω $C_2 \to \rho_2$ και μπορούμε να βγάλουμε $C_n \to \rho_n$ για την άσκηση

πολλαπλασιάζω με $\underline{x_{i-2}}$ => Γιατί με ενδιαφέρει η χρονική διαφορά (η υστέρηση των 2)

→ Ανέλιξη = οικογένεια στοχαστικών μεταβλητών (τυχαίες μεταβλητές, που δεν είναι ένα δείγμα στο χρόνο ανεξάρτητο, αλλά έχουν κάποια συσχέτιση) <u>πχ</u> παροχή ποταμού χτες και σήμερα

→ Θετική συσχέτιση = κοινός τρόπος αλλαγής => Αυτοσυσχέτιση

→ Τώρα θέλω να δημιουργήσω την 3η ροπή:

$\underline{x_i} = a \cdot \underline{x_{i-1}} + \underline{v_i} \Leftrightarrow$

$\underline{x_i'} = a \cdot \underline{x_{i-1}'} + \underline{v_i'} \Leftrightarrow$

$\underline{x_i'^3} = a^3 \cdot \underline{x_{i-1}'^3} + \underline{v_i'^3} + 3a^2 \cdot \cancel{\underline{x_{i-1}'^2} \cdot \underline{v_i}}^{0} + 3 \cdot a^2\, \cancel{\underline{x_{i-1}'} \cdot \underline{v_i'^2}}^{0} \Leftrightarrow$

$\mu_{3x} = a^3 \cdot \mu_{3x} + \mu_{3v} + \varnothing + \varnothing \Leftrightarrow \boxed{\mu_{3v} = (1-a^3)\mu_{3x}}$ 3η κεντρική ροπή (ασυμμετρία)

→ μια πιο βολική σχέση όμως είναι με τους συντελεστές ασυμμετρίας:

$C_s \cdot \sigma_v^3 = \mu_{3v}$ (εξ' ορισμού)

Συντελεστής ασυμμετρίας

Άρα από την προηγούμενη σχέση ⟺

$C_{sv} \cdot \sigma_v^3 = (1-\alpha^3) \cdot \sigma_x^3 \cdot C_{sx}$ ⟺

$C_{sv} = (1-\alpha^3) \dfrac{\sigma_x^3}{\sigma_v^3} \cdot C_{sx}$ ⟺

$C_{sv} = (1-\alpha^3) \cdot \left[\dfrac{C_0^{1/2}}{C_0^{1/2} \cdot (1-\alpha^2)^{1/2}}\right]^3 \cdot C_{sx}$ ⟺

$$\boxed{C_{sv} = (1-\alpha^3) \frac{1}{(1-\alpha^2)^{3/2}} \cdot C_{sx}}$$ Συντελεστής ασυμμετρίας

→ Τα χρησιμοποιώ όταν αντί για το ιστορικό πρέπει να φτιάξω ένα συνθετικό δείγμα κάποιων χρόνων
Προσαρμοσμένο στα ιστορικά

→ το AR(1) κρατάει μέση τιμή, τυπική απόκλιση, ασυμμετρία και συντελεστή αυτοσυσχέτισης για υστέρηση 1 (ρ_1)

Μοντέλο ARMA(1,1):

→ Για να μας βγει ακολουθούμε την σειρά βημάτων που φαίνεται στην διαφάνεια 9, του pdf "Στάσιμα Στοχαστικά Μοντέλα μιας μεταβλητής".

→ Πρώτα μ_x (μέση τιμή), μετά συνδιασπορές και μετά διασπορά C_0

Εδώ δεν είναι 0, γιατί το $\underline{x}_i$ συσχετίζεται με το $\underline{v}_i$ και το $\underline{v}_{i-1}$

$\underline{x}_i = \alpha \cdot \underline{x}_{i-1} + \underline{v}_i + b \cdot \underline{v}_{i-1}$ ⟺

(βάζουμε αναμενόμενες τιμές) ↳ σταθμισμένος όρος λευκού θορύβου σε προηγούμενη τιμή

$\mu_x = \alpha \cdot \mu_x + \mu_v + b \cdot \mu_v$ ⟺

$(1-\alpha)\mu_x = (1+b)\mu_v$ ⟺

$$\boxed{\mu_x = \frac{1+b}{1-\alpha}\mu_v}$$ Μέση τιμή

→ Πρέπει να βγάλω και την συνδιασπορά: $cov[\underline{x}_i, \underline{v}_i] = E[(\underline{x}_i - \mu_x)(\underline{v}_i - \mu_v)]$

αφαιρώ $\left.\begin{aligned} & x_i = \alpha \cdot \underline{x}_{i-1} + \underline{v}_i + b\underline{v}_{i-1} \\ (-)\ & \mu_x = \alpha \cdot \mu_x + \mu_v + b \cdot \mu_v \end{aligned}\right] \Rightarrow$

$\underline{x}_i' = \alpha \cdot \underline{x}_{i-1}' + \underline{v}_i' + b \cdot \underline{v}_{i-1}'$ ⟺ πολλαπλασιάζω με το $\underline{v}_i'$

$\underline{x}_i' \underline{v}_i' = \alpha \cdot \underbrace{\underline{x}_{i-1}' \cdot \underline{v}_i'}_{\text{Ανεξάρτητα}} + \underline{v}_i'^2 + b \cdot \underline{v}_{i-1}' \cdot \underline{v}_i'$ ⟺ περνάω αναμενόμενες τιμές στην σχέση σαν πριν $\rightarrow E[\underline{x}_i' \underline{x}_{i-1}'] = \alpha E[\underline{x}_{i-1}'^2] + E[\cancel{\underline{v}_i' \underline{x}_{i-1}'}]$

$E[\underline{x}_i' \underline{v}_i'] = \alpha \cdot E[\cancel{\underline{x}_{i-1}'}] \cdot E[\underline{v}_i'] + E[\underline{v}_i'^2] + b \cdot E[(\underline{v}_{i-1}' \cdot \underline{v}_i')] \Leftrightarrow$

$cov[\underline{x}_i, \underline{v}_i] = \phi + \sigma_v^2 + b \cdot cov[\cancel{\underline{v}_i, \underline{v}_{i-1}}] \Leftrightarrow$

Αυτοσυνδιασπορά του λευκού θορύβου με υστέρηση 1

Η αυτοσυσχέτιση του λευκού θορύβου είναι μηδέν (ανεξάρτητο μεταξύ του στον χρόνο)

Άρα καταλήγω: $\boxed{cov[\underline{x}_i, \underline{v}_i] = \sigma_v^2}$ Συνδιασπορά

→ Βγάζουμε μόνοι μας τα υπόλοιπα

Άσκηση 2β:

→ Βρίσκουμε το αρχείο BkephisosRunoff.xlsx (τα δεδομένα 100 χρόνων) → για την ακρίβεια 112 έτη

έχουμε μηνιαίες απορροές.

1α) Η ετήσια χρονοσειρά παροχών είναι το άθροισμα των μηνιαίων => είναι έτοιμη η στήλη έτος (sum οριζόντιο)

β) Οι επιμέρους μηνιαίες είναι πάλι σχηματισμένες, δηλαδή η κατακόρυφη στήλη για κάθε μήνα (12 χρονοσειρές για τους 12 μήνες)

γ) Η συνολική μηνιαία είναι όλες οι τιμές των μηνών σε χρονολογική σειρά.

↳ Ουσιαστικά σε μια στήλη όλα.

Μπορώ να το κάνω με την εντολή index (πίνακα, αριθμός σειράς, αριθμός στήλης)

Βάζω:

	i-row	j-column	Μηνιαία χρονοσειρά
	1	1	=index(εύρος πίνακα, i-row, j-columns)
κάθε φορά που αλλάζει η 12άδα προχωράω και έτος	⋮	⋮	με δολάρια
	1	12 φτιάχνω συνέχεια 12άδες	
	2	1	
	⋮	⋮	
	2	12	

μπορούμε να τα εκφράσουμε και με εντολές if

2α) Τα περιθώρια στατιστικά χαρακτηριστικά

Μέση τιμή = average του 1) μήνα, 2) έτους, 3) συνολικής μηνιαίας → Από τα παραπάνω που βγάλαμε

Τυπική απόκλιση = stdev

Συντ. Ασυμμετρίας = skew

Συντ. κύρτωσης = kurt + 3

β) Αυτοσυσχετογράμματα για την ετήσια και την συνολική μηνιαία (2) (3) → Από τα παραπάνω που βγάλαμε

→ Πρέπει να υπολογίσω την αυτοσυσχέτιση = correl

→ Άρα γράφουμε για την ετήσια:

= correl($O$10:$O$121, O11:O122)

Αυτοσυσχέτιση για υστέρηση 1 — πρέπει να αφήσει 1 τιμή εκτός

βάζω:

Υστέρηση (lag)		
1	= correl(...)	111 τιμές
2	η πάνω εντολή	110 τιμές
⋮		

→ Ουσιαστικά φτιάχνω το διάγραμμα x=lag, y=correl

→ για lag=0 => βγαίνει η χρονοσειρά με τον εαυτό της άρα το correl = 1.

Όσο ανεβάζω το lag τόσο λιγότερες τιμές μένουν για να γίνει η εκτίμηση => Άρα τόσο πιο αβέβαιο είναι

↳ Θα το πάμε μέχρι 30-40 lag. δίχως δεν έχει νόημα στο αυτοσυσχετόγραμμα το 100 lag.

→ το ίδιο πράγμα κάνουμε και για την συνολική μηνιαία. (παγώνω την 1η και παίζω με την 2η) — ουσιαστικά το correl όπως πριν

→ Η αυτοσυσχέτιση στην συνολική μηνιαία εμφανίζει μια περιοδικότητα.

↳ Παρατηρούμε: ανά 12μηνο τα θετικά peaks, ενώ ανά 6μηνο τα αρνητικά

γιατί έχουμε τα γινόμενα κάθε μήνα με τον εαυτό του στο επόμενο έτος.

↳ μεγάλη σχέση οι μήνες μεταξύ τους.

} Ένδειξη της κυκλοστασιμότητας που χαρακτηρίζει το μοντέλο που θα μελετήσουμε.

γ) Εφαρμόζουμε για κάθε μήνα το correl με lag 1 και 2.

3) Εφαρμογή του AR(1)

→ Θα πρέπει να τυποποιήσω γραμμικά (μονιμοποίηση) την χρονοσειρά μου ($\mu.\tau = 0$, $\tau.\alpha = 1$)

↳ Θα αφαιρέσω την μέση τιμή και θα διαιρέσω με την τυπική απόκλιση κάθε μήνα (με αυτές που ανήκουν σε εκείνο τον μήνα), ώστε όλοι μήνες να έχουν μέση τιμή=0 και τυπική απόκλιση=1

Τότε μπορώ να προσαρμόσω ένα μοντέλο με $\mu.\tau = 0$ και $\tau.\alpha = 1$.

→ Μετά θα πρέπει να αποτυποποιήσω γραμμικά (απομονιμοποίηση)

↳ Επαναφέρω σε κάθε μήνα την μέση τιμή και τυπική απόκλισή του

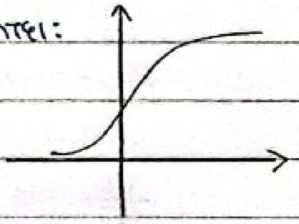
→ Ουσιαστικά θα φτιάξω έναν νέο μονιμοποιημένο πίνακα (σαν τον αρχικό)
→ για τις μονιμοποιημένες τιμές θα τις φτιάξουμε σε μηνιαία χρονοσειρά (1 στήλη με τον τύπο του index)
→ Βρίσκω μέση τιμή και τυπική απόκλιση στην μονιμοποιημένη χρονοσειρά να δούμε αν πέτυχε.

Βρίσκω το α = correl (μονιμοποιημένης, κατά 1 κάτω)
↳ στόχος είναι να διατηρεί την 1η αυτοσυσχέτιση
↓
Μπορούμε να βρούμε τα correl και για άλλα lag.

→ Φτιάχνουμε μια αθροιστική κατανομή (ή μη υπέρβασης)
= small (runoff) μονιμοποιημένη ⟶ φτιάξαμε τον άξονα x της περιθώριας κατανομής
↳ (στήλη μόνιμη με δολάρια, ΑΑ)

→ Βρίσκουμε και τις πιθανότητες που αντιστοιχούν σε αυτές τις τιμές, στην πιο μικρή: = 1/(count() + 1)
πιθανότητα μη υπέρβασης — βάζω τον ΑΑ

→ Φτιάχνω αυτό το διάγραμμα για να δω με τι μοιάζει η κατανομή και προκύπτει:



↳ σκοπός μας είναι η μέθοδος διατήρησης των ροπών (και όχι να πιάσουμε την ίδια την κατανομή)
μικρής τάξης — να μοιάζει με την ιστορική

→ η εκφώνηση λέει να χρησιμοποιήσουμε την gamma κατανομή και να πάρουμε τις παραμέτρους της. Η κατανομή gamma είναι τελείως θετικά ορισμένη, ενώ εμείς βλέπουμε ότι έχουμε και αρνητικές τιμές. (άρα πρέπει να το μετατοπίσουμε αυτό)

→ Θέλουμε να διατηρήσουμε 4 χαρακτηριστικά, τα 3 ανήκουν στην κατανομή (μέση τιμή, τυπική απόκλιση, ασυμμετρία) και την αυτοσυσχέτιση τάξης 1ης μέσω του AR(1)

→ και έχει μόνο 2 παραμέτρους η gamma κατανομή ↗ $k > 0$ ↘ $\theta > 0$ (πρόβλημα γιατί θέλουμε να διατηρήσουμε 3 χαρακτηριστικά εμείς)

→ Άρα 2 τα προβλήματά μας: 1) οι αρνητικές τιμές, 2) ότι έχουμε 2 αντί για 3 παραμέτρους.
Θέλουμε μια gamma θετικά ορισμένη => να ξεκινάει από το 0 και να είναι παράλληλη στην αρχική
⇓
Μόνο μετατόπιση θέσης. (ίδιο σχήμα)

→ Η μετατόπιση θέσης επηρεάζει την μέση τιμή μας. Διατηρούμε όμως τις άλλες ροπές (τυπική απόκλιση και ασυμμετρία)
↳ Αφαιρούν την μέση τιμή

→ Το κλασικό AR(1) είναι: $Y(i) = a \cdot Y(i-1) + V(i)$

μόνο μέσω του λευκού θορύβου μπορούμε να επηρεάσουμε την κατανομή.

Θα επιλέξουμε λευκό θόρυβο με κατανομή gamma, αφού αυτή θέλουμε να διατηρήσουμε.

Αν είχαμε προσθέσει πολλά πράγματα η κατανομή θα πήγαινε προς Gaussian, θα άσχετα με τόπους θα προσεγγίζαμε του λευκό θόρυβο

→ Πρώτα φτιάχνουμε τον λευκό θόρυβο: (*)
Βρίσκουμε k και θ (οι 2 παράμετροι της gamma κατανομής)

Στην gamma: variance $= k \cdot \theta^2$
skewness $= \frac{2}{\sqrt{k}}$

Στο AR(1):
$\mu_v = \mu_x(1-a)$
$\sigma_v^2 = C_0 \cdot (1-a^2)$
$\mu_{3v} = \mu_{3x}(1-a^3)$
Τα βρίσκουμε απευθείας από όσα ξέρουμε για την runoff

→ Άρα το $k = \frac{4}{C_v^2}$ και
$\theta = \frac{\sigma_v}{\sqrt{k}}$
Έχω βρει τα k και θ του λευκού θορύβου

→ Πριν είχαμε βρει το $C_s = \mu_3 / \sigma^3 \Rightarrow$
$C_v = S_x \cdot (1-a^3) \cdot (\sigma_x / \sigma_v)^3$

→ (*) Στην στήλη του V_i = gammainv(rand(), k, θ) και το τρέχω.
(k = α, θ = β, με \$)

→ Για να φτιάξουμε την συνθετική μας εφαρμόζουμε την εξίσωση $Y(i) = aY(i-1) + V(i)$ ξεκινάμε με αρχική τιμή την μέση τιμή του x. και μετά την εξίσωση όπως την βλέπουμε.
= \$a • το προηγούμενο y + τον λευκό θόρυβο
↳ Αυτή είναι η συνθετική μας χρονοσειρά και την πλοτάρουμε.
↓ Πρέπει να της αφαιρέσουμε το $k \cdot \theta$ για να πάρουμε $X(i)$ → έτσι θα ήταν πριν την μετατόπιση και η προσομοίωση είναι πολύ καλή
→ για να κάνουμε πιο γρήγορο έλεγχο βρίσκουμε το correl της συνθετικής χρονοσειράς και είναι ≃ ίδιο με της αρχικής. → Πιάνει το AR(1) μετά όμως αρχίζει και χάνει.
→ μπορούμε όλο αυτό να το ξαναγυρίσουμε σε πίνακα.
=[Index (βάζω στήλη κλειδωμένη, (AAy − 1)·12 + AAx)] ·τα + μ.τ
(1ο \$ στο AAy, 2ο \$ στο AAx, 2ο \$ στο τα και στο μ.τ)
$\underbrace{+\ \mu.\tau}$ για να κάνω απομονιμοποίηση

18-11-2021

Στοχαστικές Μέθοδοι
Κουτσογιάννης
Μάθημα 6ο

→ Υλίκη μας δίνει νερό και ο ταμιευτήρας του Μόρνου
↓
η βροχή στην περιοχή άρχισε να μειώνεται
→ Πρέπει να έχεις νερό να δώσεις στους Αθηναίους → συνθήκες υγιεινής
→ Πέτυχε η πρόκληση να μην μείνει ούτε μια μέρα σπίτι χωρίς νερό
→ Το μεγαλύτερο δείγμα μετρήσεων είναι στον Νείλο (ξεκίνησε από τον Θαλή)
πχ στο 620 κατασκευάστηκε ένα Νειλόμετρο ⇒ μετρούσε την στάθμη του Νείλου στην κολόνα και κρατούσαν δεδομένα. Ήταν όμως περιοχή που άλλαξε πολιτικά χέρια, αλλά οι μετρήσεις για καλή μας τύχη δεν διακόπηκαν 849 χρόνια (μέγιστη και ελάχιστη στάθμη/year)
→ Η ροζ γραμμή είναι η κλιματική γραμμή της μέσης στάθμης του Νείλου.
→ Σε milisecond μπορείς να μετρήσεις την τύρβη (μετρήσεις ροής)
→ Τα τελευταία 22 χρόνια έχει αλλάξει 0,2 βαθμούς
→ το CO_2 έχει αλλάξει τάξεις μεγέθους 400ppm (συνήθως πολύ παραπάνω)
→ η στάθμη της θάλασσας έχει πέσει 300m μέσα σε όλα τα χρόνια (τα πάντα αλλάζουν)
→ Για την υδροδότηση της Αθήνας έχει νόημα να θέσουμε ένα πλαίσιο
↓
Δουλεύουμε με στοχαστική και ελέγχουμε το σύστημα.
→ Οι φυσικές χρονοσειρές δεν συμπεριφέρονται σαν ζαριές, αλλά ομαδοποιούνται σε υγρά και ξηρά χρόνια Ομαδοποίηση (το είπε ο Hurst)
↓
Διαφοροποίηση στη φύση και στην ρουλέτα.
→ ο Hurst είναι Άγγλος υδρολόγος και μελέτησε τον Νείλο σε όλη του την ζωή. (στη φύση)
→ ο Kolmogorov περιγράφει αυτό που βρήκε ο Hurst (στα μαθηματικά) — Δεν γνωρίζονται μεταξύ τους
→ Έτσι το μοντέλο ονομάστηκε Hurst-Kolmogorov
→ Περίσσεια υδατικών πόρων αποθηκεύουμε, έλλειμμα καταναλώνουμε.
→ Αν μετρήσεις τα πάχη των δακτυλιδιών των δέντρων βρίσκεις πράγματα για το παρελθόν.

Σωρευμένη ανέλιξη: $\underline{x}_i^{(t)} := \Sigma \underline{x}_i$
↓
αθροιστικό κλιμακόγραμμα = η διασπορά της σωρευμένης ανέλιξης
→ Ορίζουμε και μέση ανέλιξη ⇝ κλιμακόγραμμα.

→ Βρίσκουμε και την αυτοσυσχέτιση

→ Το κλιμακόγραμμα και η αυτοσυσχέτιση είναι 1:1 (αν ξέρω το ένα υπολογίζω το άλλο)

στην κλίμακα 2

$$\underline{X}_1^{(2)} = \underline{x}_1 + \underline{x}_2$$

τιμή 1ου έτους → τιμή 2ου έτους

$$\underline{X}_2^{(2)} = \underline{x}_3 + \underline{x}_4$$

Διασπορά Απόδειξη: Με υπόθεση ότι η ανέλιξη είναι λευκός θόρυβος.

$$var[\underline{X}_1^{(2)}] = var[\underline{x}_1] + var[\underline{x}_2] + 2\cdot cov[\underline{x}_1 + \underline{x}_2]$$ (τύπος γενικής ισχύος)

$$\Gamma_2 = \gamma_1 + \gamma_1 + 0$$

Διασπορά — στασιμότητα — cov(λευκού) = φ ⟹ στα αθροίσματα Γ_k, στους μέσους όρους γ_k

Άρα γενικά λέμε $\boxed{\Gamma_n = k\cdot\gamma_1}$ $\boxed{\gamma_n = \frac{\gamma_1}{k}}$

→ στο μοντέλο AR(1) παίζει ρόλο και η αυτοσυσχέτιση στο κλιμακόγραμμα.
με πράξεις σε μεγάλη κλίμακα $\gamma_k = \frac{\gamma_1}{k}$ (σε μικρές κλίμακες είναι διαφορετικό)
τείνει στο $\frac{1}{k}$

→ η αλλαγή δημιουργεί αυτοσυσχέτιση. ≠ Δεν είναι μνήμη.

→ όταν φτιάχνω χρονοσειρές με αυτοσυσχετίσεις έχω υπολογίσει και την αλλαγή.

→ έχουμε πολλούς μηχανισμούς αλλαγής (τύρβη, πλημμύρα, στερεοπαροχή) που επιδράν σε διάφορες χρονικές κλίμακες.

→ Βλέπουμε το μοντέλο του Hurst με παράμετρο το H:, με $0<H<1$

Simple scaling: $\underline{X}(kt) - kt\mu =_d (k/l)^H \cdot (\underline{X}(lt) - lt\mu)$
↳ ισότητα στην κατανομή

$\boxed{\gamma_k = \frac{\gamma_1}{k^{2-2H}}}$ και $\boxed{\Gamma_k = k^{2H}\cdot\gamma_1}$ κλιμακόγραμμα

→ για $H>0{,}5$ έχει εμμονή το μοντέλο
για $H<0{,}5$ οι αλλαγές ακυρώνονται συνεχώς → αντιεμμονή

→ φτιάχνουμε 3 στήλες με AR(1) και τις προσθέτουμε. (θα το δούμε αναλυτικά σε άσκηση)

Μέθοδος συμμετρικού κυλιόμενου μέσου SMA

→ εμείς θα πιάσουμε τον Hurst-Kolmogorov αλλά μπορεί να εφαρμοστεί παντού

→ το μοντέλο δεν έχει κανένα όρο X, αλλά πολλούς όρους v

→ έχουμε και την μορφή των συντελεστών α.

Όσο πιο πολλούς όρους βάλω τόσο πιο καλά θα πιάσω το μοντέλο

→ Έστω $J=1000$ (data) και $H=0{,}85$ (data) και $H+0{,}5=1{,}35$, $\gamma_1=1$ (data), $\sqrt{\frac{(2-2H)\gamma_1}{3-2H}}=0{,}42$

n, j	a_n
-1001	με $a_n=\left(\|n+1\|^{H+0,5}+\|n-1\|^{H+0,5}-2\|n\|^{H+0,5}\right)\cdot$ Αυτό το κελί
-1000	
⋮	
1000	
1001	

περιμένω ότι θα είναι ένα συμμετρικό σχήμα

→ όσο αυξάνω το H απομακρύνομαι από το 0

v_j	X_t
=normsinv(rand()) ↓ 3000 σύνολο ↙↘ 2000 + 1000	=sum(όλα τα a_n · όλα τα v_j) και μετά (ctrlshift↓) χωρίς $

→ Θα δέσουμε 1% αστοχία στα 1000χρόνια παραγωγή δεδομένων.

→ Πρέπει όσα είναι τα α προσθέτω 1000 ακόμα v_j

→ μπορούμε να βγάλουμε και καμια 50αριά αυτοσυσχετίσεις της στήλης X_t

↳ πλοτάρω τις αυτοσυσχετίσεις και πάει αρκετά καλά σαν Hurst-Kolmogorov

↓ μπορούμε να το κάνουμε και λογαριθμικό ⇒ ευθεία πλοτάρεται η σχέση δύναμης στο log

↳ Δεν θα είναι καλή η ευθεία για 1000 μόνο δεδομένα. θέλει 10.000

25-11-2021

Στοχαστικές Μέθοδοι
Άννυ
Μάθημα 7ο

Άσκηση 3:
pdf: Εμμονή και ανελίξεις απλής ομοιοθεσίας (Hurst-Kolmogorov)
→ κλίμακα είναι χρονικό παράθυρο που είτε συναθροίζω, είτε μέση τιμή των ετών αυτών.
→ αναμένουμε ότι η μέση τιμή εξομαλύνει.
↳ το AR(1) δείχνει πολύ καλύτερα αποτελέσματα.
→ για να εκτιμήσουμε το φαινόμενο της αλλαγής χρησιμοποιούμε το κλιμακόγραμμα.
(εκτιμώ τις διασπορές) — συναθροισμένο → sum; απλό → μ.τ.
↳ ποσοτικοποιεί τις διαφορές την διασπορά AR(1) ή λευκού θορύβου ⇒ όσο πιο έντονη η διαφορά τόσο

Ερώτημα 1ο μεγαλύτερη η δυναμική του φαινομένου Hurst-Kolmogorov
→ πάλι δουλεύουμε με τον Βοιωτικό Κηφισσό.
→ υπάρχει θεωρητική σχέση για το κλιμακόγραμμα (για Hurst, AR(1) και λευκού θορύβου)
↳ Διαφορετικοί τύποι
→ παίρνουμε την τελευταία χρονοσειρά (την ετήσια χρονοσειρά)
Βάζω κάπου την κλίμακα k = 1 − 11 (για πολλές κλίμακες)
1: ίδια η χρονοσειρά copy paste

k=2 : 1ο κελί = μέση τιμή (1ου, 2ου)
2ο κελί = μέση τιμή (3ου, 4ου)
⋮ κτλπ.

Στην στήλη k βάζω έναν αύξοντα αριθμό (μετρητής)
Θα φτιάξουμε έναν κώδικα = offset (1ο κελί κ=1 $, κλίμακα 2·(αα1 −1), 0, κλίμακα 2)
και απ' έξω βάζω average

→ και το τραβάμε δεξιά μέχρι την κλίμα 11. (όσο μεγαλώνω κλίμακα τόσο λιγότερα νούμερα παίρνω)
μετά θέλω να εκτιμήσω διασπορά, άρα δεν έχει νόημα να αυξήσω την κλίμακα πολύ ⇒ χρειάζομαι τουλάχιστον 10 στοιχεία.

→ πάνω απ'την υλίμαυα μετράω: floor (112 / υλίμαυα)
στρογγυλοποίηση προς τα υάτω — με $
→ μπορούμε να πάμε την προηγούμενη σχέση = if(στήλη αα με 1ο $ <= του floor, average...., "")
"" → βάλε μουνευό.
→ Έτσι βγάλαμε τις χρονοσειρές σε υλίμαυα
→ Για το υλιμαυόγραμμα βρίσυω το var σε υάθε μια υλίμαυα => πως αλλάζει η διασπορά σε υλίμαυα
Αυτό είναι το εμπειριυό υλιμαυόγραμμα.
υαι υάνω το διάγραμμα: x = υλίμαυες
y = διασπορές
μπορώ να βάλω υαι λογαριθμιυούς άξονες

Θεωρητιυό υλιμαυόγραμμα του AR(1):
→ βρίσυω το $\rho_1 = 0{,}365 = r$ (του τύπου) υαι εφαρμόζω τον τύπο
$g_1 = var(\text{ιστοριυής}) = 25719{,}98$

Λευυού θορύβου: => εφαρμόζω τον τύπο.

Hurst:
$\ln \gamma_k$ τα βρίσυω οριζόντια
$\ln k$
υλίμαυας
ή θα μπορούσα να ζητήσουμε ένα διάγραμμα με σχέση δύναμης υαι να πάρω τον εκθέτη
=slope(lnγk, lnk)
υλίση — όλων των μελιών
" 2H-2 (ουσιαστιυά βάζω ln στην σχέση)
H=(υλίση+2)/2 = 0,8

H.K.	= τύποι θεωρητιυοί
AR(1)	για το γk τους.
W.N.	(υαι τραβάμε δεξιά)

→ μετά βλέπω πως προσαρμόζονται στο εμπειριυό → τα βάζω όλα στο ίδιο διάγραμμα.
↳ έλεγχος υαταλληλότητας
→ το H.K προσαρμόζεται υαλύτερα.
→ για H=0,5 έχουμε το μοντέλο του λευυού θορύβου.

	Λευκός Θόρυβος	AR(1)	H. k
→ Εμπειρική αυτοσυνδιασπορά Θεωρητική —ιι— τύποι	∅ (εξ' ορισμού) και η αυτοσυσχέτιση και η αυτοσυνδιασπορά	$c_n = a^n \cdot c_0$ ρ	Είναι ανεξάρτητη της κλίμακας. Ισχύει για όλες τις κλίμακες.

→ Βρίσκω το covar της ιστορικής (εμπειρική) για lag μέχρι 30

όμοια με correl

Βρίσκω αυτοσυνδιασπορά

Αυτοσυσχέτιση:

lag	AR(1)
1	$= \rho$
2	$= \rho^{\wedge 2}$
⋮	$= \rho^{\wedge 3}$
15	

$= \gamma_1 \cdot r_n$ = αυτοσυνδιασπορά (αυτοσυσχέτιση, τύπος).

→ Δεν βάζω απόλυτη τιμή, γιατί έχω θετικά lag.

→ και κάνω πάλι τα διαγράμματα.
Εμπειρικό, AR(1), Hk (με το lag στον x)
(lag, αυτοσυνδιασπορά)

↳ πλοτάρουμε και βλέπουμε ότι φθίνει ευθέτικά και μηδενίζει

→ Παρατηρούμε ότι το Hk προσαρμόζεται καλύτερα πάλι στο εμπειρικό.

Αυτοσυνδιασπορά:

$= \rho^{\wedge lag} \cdot \gamma_1$

↳ της ιστορικής (που βρήκα πριν)

Άλλος ένας τρόπος να ποσοτικοποιήσουμε την αλλαγή, αλλά το κλιμακόγραμμα είναι πιο καλή η προσαρμογή. (οι 2 τρόποι είναι ισοδύναμοι θεωρητικά για να δούμε την αλλαγή)

Ερώτημα 2ε:

→ του AR(1) όπως κάναμε στην προηγούμενη άσκηση. (με κανονική κατανομή)
→ του H.k με 2 τρόπους → 3 AR(1) → πρόσθεση
↳ SMA → μοντέλο συμμετρικού κυλιόμενου μέσου.

SMA:

→ σαν να παίρνεις σταθμισμένους μέσους Λ.Θ. και να τους προσθέτεις
↳ φτιάχνεις κάτι που θα έχει αυτοσυσχέτιση.

$\gamma_1 = 25719{,}98$
$H = 0{,}8$
} τα παίρνω από πριν

→ φτιάχνω την χρονοσειρά των βαρών.
J = μετρητής βάρους και είναι συμμετρικοί. (-1000 → 1000)

Άρα 2001 βάρη. ⇒ καλύτερη ακρίβεια

→ φτιάχνω σε ένα κελί το $= \frac{\sqrt{(2-2H)\cdot \gamma_1}}{3-2H} = 72,45$

j	a_j	v (χρονοσειρά Ν.Θ.)
=-1000	$=72,45\cdot\left(\lvert n+1\rvert^{H+0,5}+\lvert n-1\rvert^{H+0,5}-2\lvert n\rvert^{H+0,5}\right)$ εδώ βάζω ABS	↗ Η διασπορά θα είναι 1. Βάζω αναμενόμενες τιμές.
-999	με $n=j$	$E[x_t]=E[a_j v_{t-j}+\ldots+a_j v_{t+j}]$
⋮		$\mu_x = a_j\cdot E[v_{t-j}]+\ldots+a_j[v_{t+j}] =$
		$=\mu_v(a_j+\ldots+a_j) \Leftrightarrow$
+1000		$\mu_v = \frac{\mu_x}{\sum_{-j}^{j} a_j} = 0,2465$

→ Βρίσκω και μ_x της ιστορικής $=368,9$

Άρα $v = \text{norminv}(\text{rand}(), \mu_x, 1)$

Το τρέχουμε μέχρι 4000.

SMA

$=\text{sum}(\underset{\text{όλων}}{a_j}\cdot\underset{\text{όλων}}{V})$ χωρίς \$ και πατάω ctrl shift enter

και φτάνουμε μέχρι το 1000.

Βρίσκουμε μ_x και var του SMA να είναι κοντά.

3 AR(1):

ρ, φ, ξ = συντελεστές αυτοσυσχέτισης για τα 3 AR(1)

→ η αυτοσυσχέτιση του αθροίσματος = με αυτή του μοντέλου

	A	B	Γ	
	ρ	φ	ξ	=> Αυτοσυσχετίσεις
	$=1,52(H-0,5)^{1,32}$	$=0,953-7,69\cdot(1-H)^{3,85}$	$=0,993+0,007\cdot H$ (για $H>0,76$)	
m_v	Φ	Φ	Φ	
var_x	$=(1-c_1-c_2)\gamma_1$	$=c_1\cdot\gamma_1$	$=c_2\cdot\gamma_1$	
var_v	$=(1-\rho^2)\cdot$var_x	$=(1-\varphi^2)$var_x	$=(1-\xi^2)$var_x	
τυπική απόκλιση f_v	$=$ var_x^{0,5}	όμοια	όμοια	

r_Hk Θεωρητική αυτοσυσχέτιση του Hk

1: $=0,5\left[(1+1)^{2H}+(1-1)^{2H}\right]-1^{2H}$ => Λύνω σύστημα 2 αγνώστων το c_1 και c_2

100: = όμοιο με το 100

→ Φέρνουμε το σύστημα στην μορφή $A \cdot x = B \Rightarrow x = A^{-1} \cdot B$

$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$

Για $n=1$: $r_{HK_1} = (1 - C_1 - C_2)\rho^1 + C_1 \cdot \varphi^1 + C_2 \cdot \mathcal{7}^1$ = το ερήμα πριν

Για $n=100$ $r_{HK_{100}} = (1 - C_1 - C_2 \cdot \rho)^{100} + C_1 \cdot \varphi^{100} + C_2 \cdot \mathcal{7}^{100}$ = το ερήμα πριν

Μορφή 1: $C_1(\varphi - \rho) + C_2 \cdot (\mathcal{7} - \rho) = r_{HK_1} - \rho$

Μορφή 100: $C_1(\varphi^{100} - \rho^{100}) + C_2 \cdot (\mathcal{7}^{100} - \rho^{100}) = r_{HK_{100}} - \rho^{100}$

σε μορφή μητρώου: $\begin{bmatrix} \varphi - \rho & \mathcal{7} - \rho \\ \varphi^{100} - \rho^{100} & \mathcal{7}^{100} - \rho^{100} \end{bmatrix} \cdot \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} r_{HK_1} - \rho \\ r_{HK_{100}} - \rho^{100} \end{bmatrix}$

→ minverse A $= A^{-1}$ με ctrl shift enter. (για να βρω τον ανάστροφο)

→ διαδικασία επίλυσης σε μορφή πινάκων.

→ πολλαπλασιασμός πινάκων mmult( )

→ τώρα συμπληρώνω τα προηγούμενα

→ πάμε να κάνουμε σύνθεση του μοντέλου.

ίση με την μ.τ των AR(1)

0

το κάνω x 3 φορές και μετά παίρνω sum.

1 $= \rho \cdot$ από πάνω + norminv(rand(), 0, st-v)

2

3

:

1000

διατηρεί:
1) διασπορά
2) Hurst
3) Αυτοσυσχετίσεις

⇓

πρέπει να διατηρεί και την μέση τιμή. (της ιστορικής)

x (διορθωμένη)

κελία
= sum − μ.τ (που έχει) + μ.τ (ιστορικής) → ελέγχω μ.τ και var

⋮

(+) τι Hurst βγάζει η χρονοσειρά που φτιάξαμε (copy το 1ο φύλλο και βλέπουμε)

→ κάνουμε τοίδιο πράγμα και για Markov

Scanned with CamScanner