21-6-2022

uoa.webex.com/meet/aburnetas

Bootstrap - simulation

$$X \sim F(\theta)$$

Θέλουμε να εκτιμήσουμε το θ.

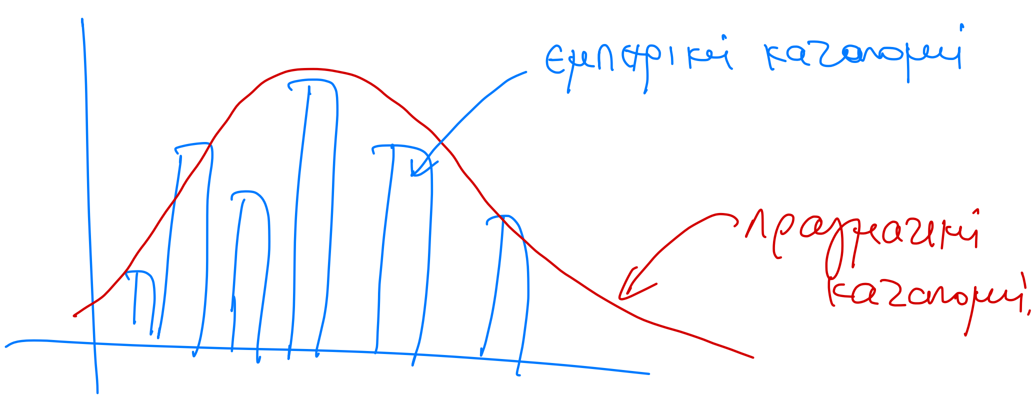
Δείγμα (πραγματικό από τον πληθυσμό)

$(x_1, x_2, \dots, x_n)$ ανεξ $\sim F$

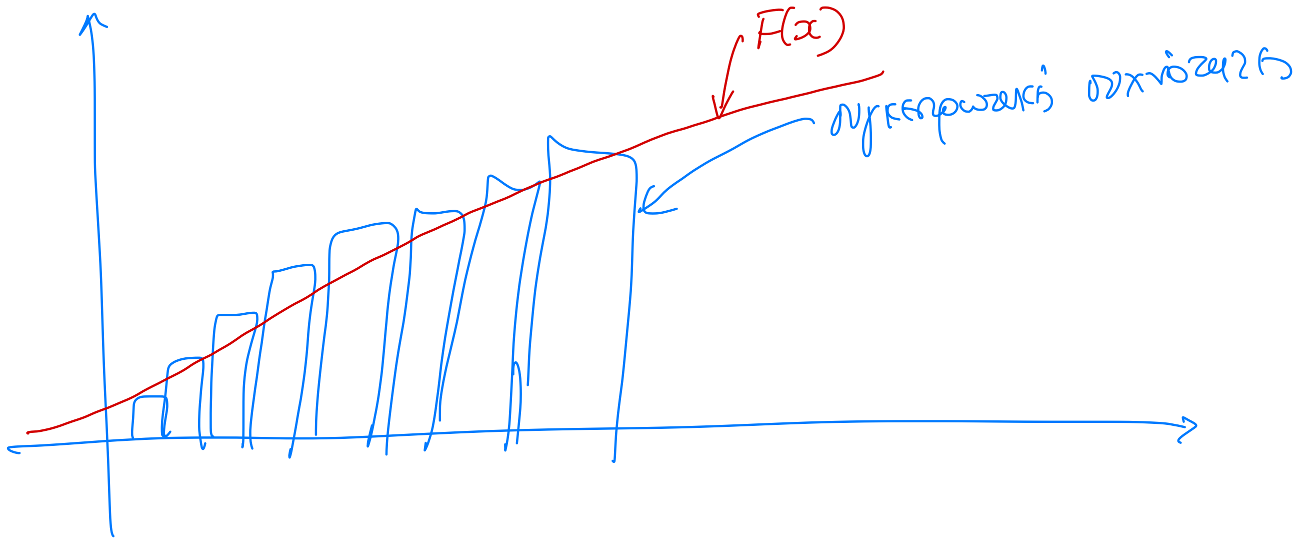
$$\hat{\theta} = \underbrace{T(x_1, \dots, x_n)}_{\text{sampling distribution}} \leftarrow$$ στατιστική συνάρτηση εκτιμήτρια του θ. (σημειακή εκτίμηση)

→ τυχαία μεταβλητή λόγω τυχαίας δειγματοληψίας

Ιστόγραμμα



αν n μεγάλο ιστόγραμμα → σππ



Βασική Ιδέα 1: Γεννήτρια τυχαίων αριθμών από εμπειρική κατανομή.

Η εμπειρική κατανομή διακριτή

Π.χ.

$n = 10$

$x = (2, 3, 2, 5, 6, 1, 3, 2, 3, 4)$

συχν.

1 : 1/10   5 : 1/10

2 : 3/10   6 : 1/10

3 : 3/10

4 : 1/5

Εμπειρική κατανομή

Διακριτή	1	2	3	4	5	6
p	1/10	3/10	3/10	1/10	1/10	1/10

Βασική Ιδέα 2: Αν πάρω ένα δείγμα μεγέθους n από το αρχικό μου δείγμα (με επανάθεση)

⟹ η κατανομή αυτού του νέου δείγματος είναι η εμπειρική κατανομή του αρχικού.

R: sample(x, i, replace=TRUE)

↳ τυχαίο σύνολο (δείγμα) μεγέθους i από τα στοιχεία του x, με επανάθεση (επιτρέπονται οι επαναλήψεις)

Πραγματικότητα

Άγνωστη κατανομή F

Ένα δείγμα ~~από~~ (n) από F $(\underline{x})$

θ : άγνωστη παραμ.

$\hat{\theta} = T(\underline{x})$: εκτίμηση

Αν $n \to \infty$ $\hat{\theta} \to \theta$ (συνέπεια)

Δείγμα $(\underline{x})$

F_e : εμπειρική (από $\underline{x}$)

$\hat{\theta}$ (αληθινή τιμή παραμέτρου κάτω από F_e)

Μέσω bootstrap (sample) παίρνω N δείγματα μεγέθους $\underline{n}$ από F_e

Σε καθένα υπολογίζω

T_1
T_2
$\vdots$
T_N

διαφορετικές τιμές της εκτιμήτριας T.

Π.Χ. bias $\overline{T}$

$E(T) - \theta = b$

$\Longrightarrow$ $\widehat{\text{bias}}$

$$\boxed{\overline{T_N} - \hat{\theta}} = \hat{b}$$

Εκτίμηση bias μέσω bootstrap.

Έχουμε : $\hat{\theta}$: εκτίμηση του θ

$\hat{b}$: εκτίμηση του bias b

$$\hat{b} = \widehat{ET - \theta}$$

Διόρθωση του $\hat{\theta}$ για να λάβουμε υπόψη το bias

Βελτιωμένη εκτίμηση του θ

$$\tilde{\theta} = \hat{\theta} - \hat{b} = \hat{\theta} - (\overline{T}_N - \hat{\theta})$$

$$\Rightarrow \boxed{\tilde{\theta} = 2\hat{\theta} - \overline{T}_N}$$

bias correction μέσω bootstrap.

Παράδειγμα 2

Έστω ότι θέλουμε να εκτιμήσουμε το μέσο τετραγωνικό σφάλμα του T.

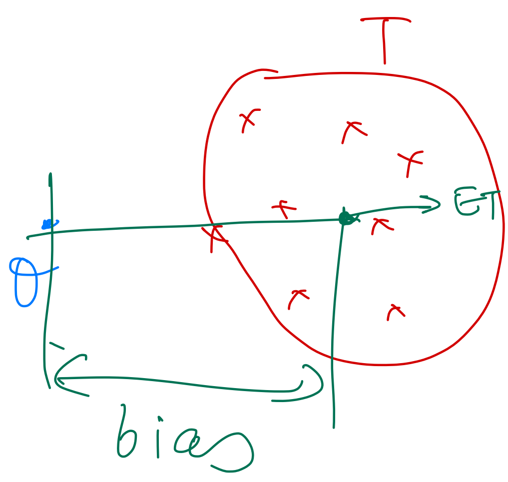
$$MSE = E(T-\theta)^2$$

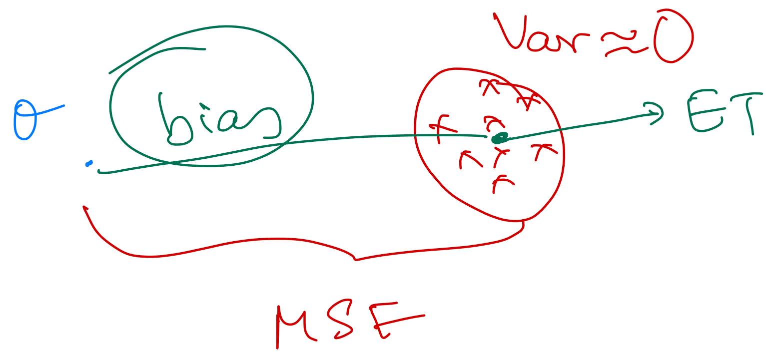
Αν η T αμερόληπτη $E(T)=\theta \Rightarrow$

$$\Rightarrow MSE = E(T-E(T))^2 = Var(T)$$

$$\widehat{MSE} = s^2$$ (δειγματική διασπορά)

Αν T δεν είναι αμερόληπτη $MSE = ?$





Συνήθως: $MSE = Var(T) + (ET-\theta)^2$

bias-variance decomposition

Εκτίμηση MSE μέσω **bootstrap**

$T_1, \dots, T_N$ από N δείγματα bootstrap.

$$\widehat{MSE} = \frac{1}{N}\sum_{j=1}^{N}\left(T_j-\hat{\theta}\right)^2$$

Παράδειγμα

Έστω $X \sim Exp(\lambda)$

Δείγμα μεγ. n: $x_1, \ldots, x_n$

$$\hat{\lambda} = \frac{1}{\bar{X}_n}$$

Έστω λ άγνωστο (συγκεκριμένο προέρχεται από συγκεκρ. πληθυσμό)

Εκτίμηση bias μέσω bootstrap.

1. N bootstrap δείγματα από X (N δείγματα μεγ. n με επανάθεση)

   Σε κάθε δείγμα $j = 1, \ldots, N$

   υπολ. $T_j = \dfrac{1}{(\bar{X}_n)_j}$

2. $$\bar{T}_N = \frac{1}{N} \sum_{j=1}^{N} T_j$$

3. $$\hat{b} = \bar{T}_N - \frac{1}{\bar{X}_n}$$ ← (αρχικό)

④ Corrected estimate (λ) : $\boxed{\tilde{\lambda} = \frac{2}{\bar{X}_n} - \bar{T}_N}$

$= \hat{\lambda} - \hat{b}$

Markov Chain Monte Carlo (MCMC)

Αλγόριθμος Metropolis-Hastings

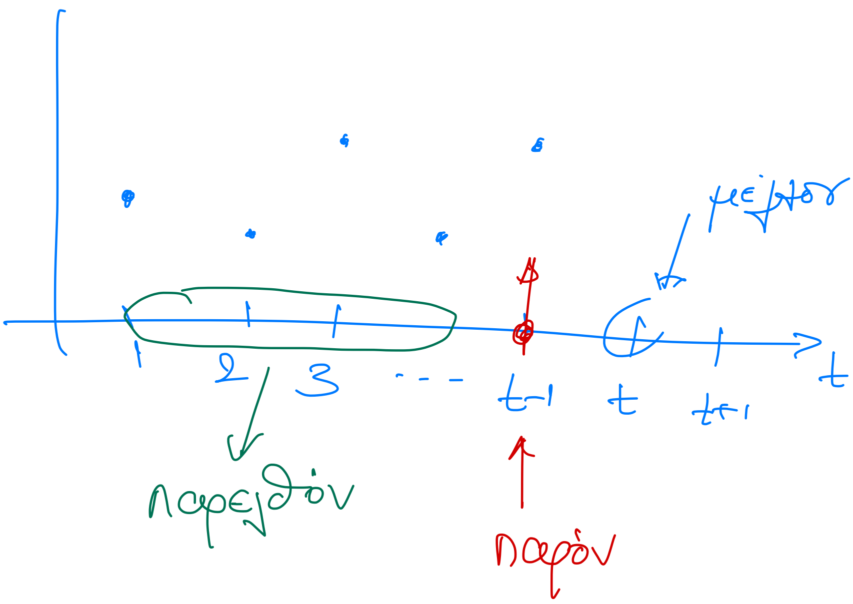
Μαρκοβιανή διαδικασία

Χρονοσειρά $(X_1, X_2, \dots, X_t, X_{t+1}, \dots)$

X_t: τιμή ενός μεγέθους (κατάσταση συστήματος) κατά την περίοδο t.

Η κατανομή της X_t δεδομένων $\underbrace{X_1, X_2, \dots, X_{t-1}}$ εξαρτάται μόνο από το X_{t-1} κι' όχι από τις προηγούμενες τιμές.

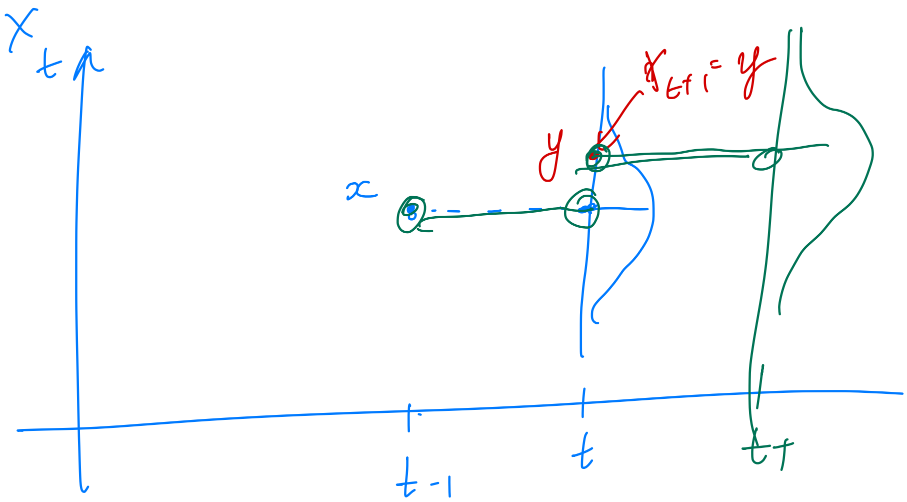
Μαρκοβιανή ιδιότητα



Συνάρτηση πιθανότητας μετάβασης (transition kernel)

$$X^{(t)} \,\Big|\, X^{(t-1)} = x$$ τυχαία μεταβλητή με σππ $q(y|x)$

π.χ. δεδ. $X^{(t-1)} = x$: $X^{(t)} \sim N(x, 1)$

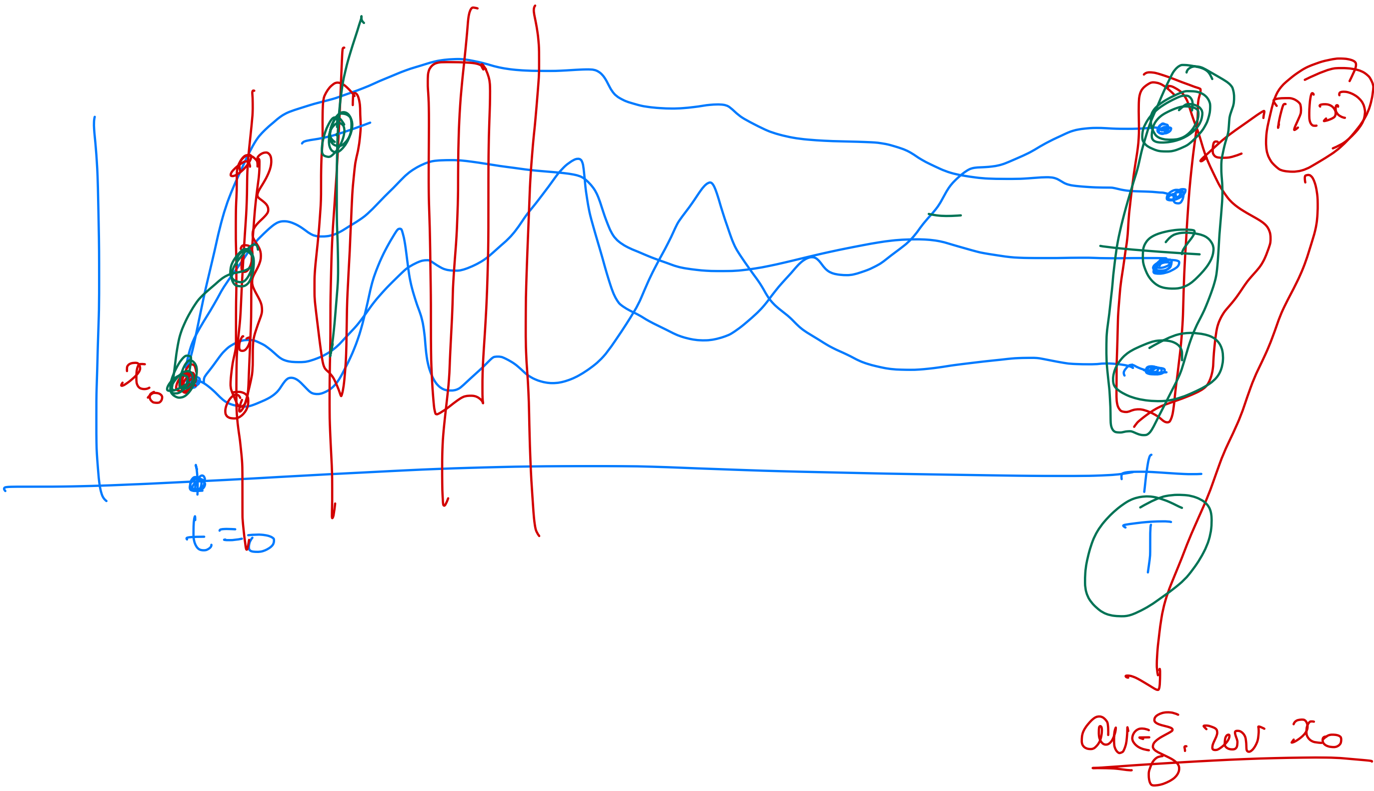


Στάσιμη και Οριακή Κατανομή.

Κάτω από κάποιες γενικές υποθέσεις

η X_t όταν $t \to \infty$ έχει οριακή κατανομή $\pi(x)$.

$$P(X_t = x) \xrightarrow[t\to\infty]{} \pi(x)$$



Η $\pi(x)$ προκύπτει από $q(y|x)$ μέσω κατάλληλων εξισώσεων.

Βασική Ιδέα 1:

Έστω $f(x)$ μια κατανομή για την οποία θέλουμε να κατασκευάσουμε γεννήτρια.

Αν βρούμε ?? ένα transition kernel $q(y|x)$ τέτοιο ώστε $\pi(x) = f(x)$

αυτό δίνει μια γεννήτρια:

Αλγόριθμος

$X_0 : x$ αυθαίρετο

$X_1 \sim q(y|x_0)$ (από προσομοίωση)

$X_1 = x_1 : X_2 \sim q(y|x_1)$

$X_3 \sim q(y|x_2)$

Βασική Ιδέα 2: Πως προσδιορίζουμε το κατάλληλο kernel $q(y|x)$?

Αλγόριθμος Metropolis-Hastings

Εστω οποιοδήποτε kernel $q(y|x)$

1. Δημιουργούμε $Y \sim q(y|x_{t-1})$ εστω $Y = y$, $X_{t-1} = x$ (προσομοίωση)

2. Εστω $\rho(x,y) = \min\left\{ \dfrac{f(y)\, q(x|y)}{f(x)\, q(y|x)},\ 1 \right\}$

$$0 \le \rho(x,y) \le 1.$$

3. <u>Κρατάμε</u> το y δηλ. $X_t = y$ με πιθ. p

Το <u>απορρίπτουμε</u> δηλ. $X_t = x$ με πιθ. $1-p$

(μέσω προσομοίωσης

$U \sim U(0,1)$ → αν $U \leq p \Rightarrow X_t = y$

→ $U > p \Rightarrow X_t = x$