# Στοχαστική Ανάλυση Χρονοσειρών

**Περιεχόμενο:**

**Εκτιμητές Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimators – MLE)**

**Πρόβλεψη**

**Αξιολόγηση Προβλέψεων**

Επαμεινώνδας Διαμαντόπουλος
Πέμπτη 25 Απριλίου 2024

# Εκτιμητές μέγιστης πιθανοφάνειας

# Εκτιμητές μέγιστης πιθανοφάνειας

Στην Στατιστική, η εκτίμηση μέγιστης πιθανότητας (MLE) είναι μια μέθοδος εκτίμησης των παραμέτρων μιας άγνωστης κατανομής πιθανότητας, από ένα σύνολο παρατηρούμενων δεδομένων.

Αυτό επιτυγχάνεται με τη μεγιστοποίηση μιας συνάρτησης πιθανότητας, έτσι ώστε, σύμφωνα με το υποτιθέμενο στατιστικό μοντέλο, τα παρατηρούμενα δεδομένα να είναι πιο πιθανά να εμφανιστούν.

Το σημείο στο χώρο των παραμέτρων που μεγιστοποιεί τη συνάρτηση πιθανότητας ονομάζεται εκτιμητής μέγιστης πιθανότητας.

Η διαδικασία είναι εφικτή τόσο για συνεχείς όσο και για διακριτές τμ. Η μέθοδος προτάθηκε από τον Fisher το 1912.

# Εκτιμητές μέγιστης πιθανοφάνειας

Έστω ότι τα δεδομένα $x = (x_1, \ldots, x_n)$, προέρχονται από μία κατανομή της οποίας γνωρίζουμε το είδος της κατανομής αλλά όχι τις παραμέτρους αυτής. Έστω επίσης πως οι παρατηρήσεις είναι ανεξάρτητες.

Ορίζουμε τη συνάρτηση πιθανοφάνειας L ως εξής :

$$L(\theta) = L(X, \theta) = f(x_1, \ldots, x_n | \theta)$$

Αναζητούμε την τιμή της παραμέτρου θ που μεγιστοποιεί τη συνάρτηση L. Η τιμή που βρίσκουμε με την παραπάνω διαδικασία ονομάζεται

**Εκτιμητής Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation (MLE)).**

# Εκτιμητές μέγιστης πιθανοφάνειας

Οι παρατηρήσεις θεωρούνται ανεξάρτητες, άρα

$$f(x_1, \ldots, x_n|\theta) = f(x_1|\theta)\cdot f(x_2|\theta)\cdot\ldots\cdot f(x_n|\theta)$$

Η συνάρτηση πιθανοφάνειας L γράφεται:

$$L(\theta) = L(X, \theta) = f(x_1|\theta)\cdot f(x_2|\theta)\cdot\ldots\cdot f(x_n|\theta)$$

Για να βρούμε την τιμή της παραμέτρου θ που μεγιστοποιεί τη συνάρτηση L, αρκεί να βρούμε την τιμή που μεγιστοποιεί την

$$l(\theta) = \ln(L(\theta)) = \Sigma\ \ln(f(x_1|\theta)).$$

Ο **εκτιμητής μέγιστης πιθανοφάνειας** προκύπτει από την επίλυση της εξίσωσης

$$d/d\theta\ l(\theta) = 0$$

# Εκτιμητές μέγιστης πιθανοφάνειας

**Άσκηση 1.** Έστω X μία ΤΜ με κατανομή

| x | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f_X(x)$ | $2\theta/3$ | $\theta/3$ | $2(1 - \theta)/3$ | $(1 - \theta)/3$ |

Για να εκτιμήσουμε την άγνωστη παράμετρο θ, υλοποιούμε δειγματοληψία 10 ανεξάρτητων τιμών και παίρνουμε το δείγμα (3, 0, 2, 1, 3, 2, 1, 0, 2, 1). Να βρεθεί η τιμή της παραμέτρου θ με τη μέθοδο του Εκτιμητή Μέγιστης Πιθανοφάνειας.

# Εκτιμητές μέγιστης πιθανοφάνειας

**Λύση Άσκησης 1.**

$L(X, \theta) = f(\,(3, 0, 2, 1, 3, 2, 1, 0, 2, 1) \mid \theta)$

$= f(3|\theta) \cdot f(0|\theta) \cdot f(2|\theta) \cdot f(1|\theta) \cdot f(3|\theta) \cdot f(2|\theta) \cdot f(1|\theta) \cdot f(0|\theta) \cdot f(2|\theta) \cdot f(1|\theta)$

$= [f(0|\theta)]^2 \cdot [f(1|\theta)]^3 \cdot [f(2|\theta)]^3 \cdot [f(3|\theta)]^2$

$= [2\theta/3]^2 \cdot [\theta/3]^3 \cdot [2(1 – \theta)/3]^3 \cdot [(1 – \theta)/3]^2$

$= \theta^5 \cdot (1 – \theta)^5 \cdot 2^5 / 3^{10}$.

$l(\theta) = \ln L(X, \theta) = 5 \ln \theta + 5 \ln(1 – \theta) + c$.

Η συνάρτηση $l(\theta)$ μεγιστοποιείται για $\theta = ½$. Αυτή είναι η τιμή του εκτιμητή μέγιστης πιθανοφάνειας. Δηλαδή, η τιμή $\theta = ½$ είναι αυτή που μεγιστοποιεί την πιθανότητα το δείγμα που έχουμε να έχει προέλθει από την υποτιθέμενη κατανομή.

# Εκτιμητές μέγιστης πιθανοφάνειας

**Άσκηση 2.** Έστω ότι η τ.μ. Χ μετράει το χρόνο μεταξύ αφίξεων σε μία διαδικασία Poisson και πως Χ ~ Exp(λ). Παρακολουθούμε δείγμα 4 τιμών της Χ και βρίσκουμε τις τιμές 1.23, 3.32, 1.98, 2.12. Να βρεθεί ο Εκτιμητής Μέγιστης Πιθανοφάνειας της παραμέτρου λ.

**Λύση Άσκησης 2**

Από την υπόθεση έχουμε πως Χ ~ Exp(λ), δηλαδή ότι $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.

Ορίζουμε $L(X, \theta) = f(\,(1.23,\ 3.32,\ 1.98,\ 2.12)\mid \lambda)$

$$= f(1.23|\lambda) \cdot f(3.32|\lambda) \cdot f(1.98|\lambda) \cdot f(2.12|\lambda)$$

$$= \lambda e^{-1.23\lambda} \cdot \lambda e^{-3.32\lambda} \cdot \lambda e^{-1.98\lambda} \cdot \lambda e^{-2.12\lambda}$$

$$= \lambda^4\, e^{-8.65\lambda}\ .$$

Είναι $l(\theta) = \ln L(X, \theta) = 4 \ln \lambda - 8.65\,\lambda$.

Η συνάρτηση $l(\theta)$ μεγιστοποιείται για $\lambda = 0.46$, τιμή που μεγιστοποιεί την πιθανότητα το δείγμα που έχουμε να έχει προέλθει από την κατανομή Exp(λ).

# Αμερόληπτοι και συνεπείς εκτιμητές

Πόσο καλές είναι οι εκτιμήσεις των παραμέτρων που προκύπτουν από τους τύπους που χρησιμοποιούμε;

Στην Στατιστική είναι επιθυμητό ο εκτιμητής $\theta_n$ να είναι

- **αμερόληπτος** (unbiased)

$$\mathrm{Bias}(\theta_n, \theta) = E_{x|\theta}(\theta_n - \theta) = 0.$$

(ανεξάρτητα από το μέγεθος του δείγματος, ο εκτιμητής δίνει τη σωστή τιμή της παραμέτρου ως αναμενόμενη τιμή)

- **συνεπής** (consistent)

$$\mathrm{plim}_{n\to\infty}\theta_n = \lim_{n\to\infty} P(|\theta_n - \theta| > \varepsilon) = 0, \text{ για κάθε } \varepsilon > 0.$$

(η αύξηση του μεγέθους του δείγματος οδηγεί σε καλύτερη πρόβλεψη της παραμέτρου)

Και οι δύο έννοιες αποκτούν νόημα με την προϋπόθεση πως γνωρίζουμε το είδος της κατανομής που ακολουθεί ο πληθυσμός από τον οποίο πήραμε το δείγμα.

# Μέγιστη Πιθανοφάνεια $\not\rightarrow$ Αμεροληψία

**Παράδειγμα 1 (με αφορμή την άσκηση 2)**
Έστω X μία μεταβλητή που γνωρίζουμε ότι ακολουθεί την Exp(λ) αλλά δεν γνωρίζουμε το λ (= πλήθος αφίξεων / μονάδα χρόνου). Για το λόγο αυτό παρατηρούμε n διαφορετικές διάρκειες μεταξύ αφίξεων $X_1$, $X_2$, …, $X_n$ και εκτιμούμε τον ρυθμό λ από τον τύπο:

$$\hat{\lambda} = \frac{n}{X_1 + X_2 + \ldots + X_n}$$

Ο εκτιμητής αυτός **δεν είναι αμερόληπτος**. Πράγματι, για n = 1, είναι

$$E(\hat{\lambda}) = E\left(\frac{1}{X_1}\right) = \int_0^{+\infty} \frac{\lambda}{x} e^{-\lambda/x} dx = +\infty$$

ενώ για n > 1 είναι $X_1 + X_2 + \ldots + X_n$ ~ Γάμμα(n, λ), άρα $1/(X_1 + X_2 + \ldots + X_n)$ ~ Inverse Γάμμα(n, λ) και

$$E(\hat{\lambda}) = E\left(\frac{n}{X_1 + X_2 + \ldots + X_n}\right) = n\frac{\lambda}{n-1} \neq \lambda .$$

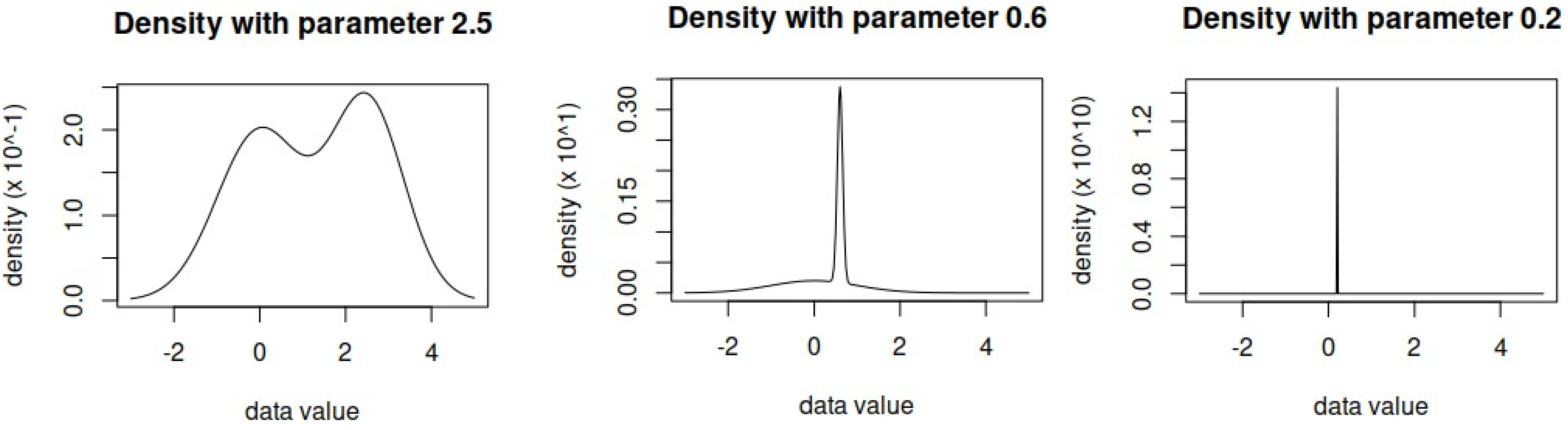
Ιδιαίτερα, συμπεραίνουμε, ότι ο $\hat{\lambda} = \dfrac{n-1}{X_1 + X_2 + \ldots + X_n}$ είναι ένας αμερόληπτος εκτιμητής του λ.

Σημείωση: Αν X ~ Inverse Γαμμα(α, β) τότε $E(X^k) = \beta^k \, \Gamma(\alpha - k) / \Gamma(\alpha)$.

# Μέγιστη Πιθανοφάνεια $\not\longrightarrow$ Συνέπεια

**Παράδειγμα 2**

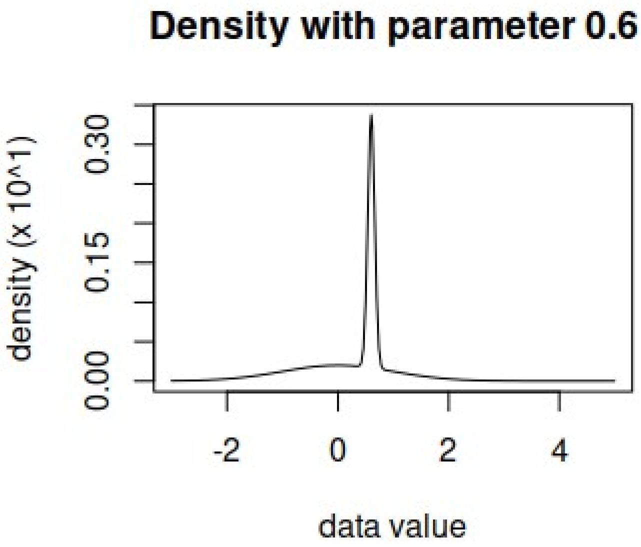
Έστω X τμ για την οποία γνωρίζουμε ότι $X \sim 1/2N(0,1) + 1/2N\left(t, e^{-\frac{2}{t^2}}\right)$

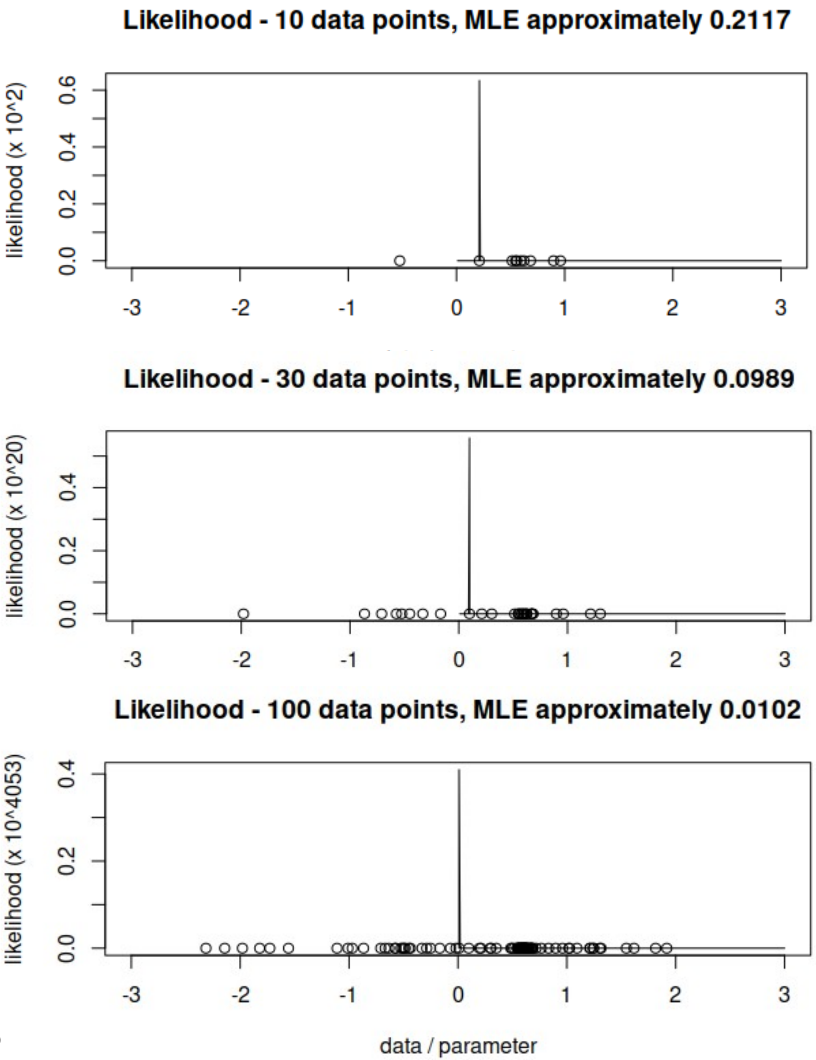



Πηγή: https://radfordneal.wordpress.com/2008/08/09/inconsistent-maximum-likelihood-estimation-an-ordinary-example/

# Μέγιστη Πιθανοφάνεια $\not\longrightarrow$ Συνέπεια

**Παράδειγμα 2**
Για t = 0,6, και $\{x_i\}_{i \leq n}$, n = 10 ή 30 ή 100 παρατηρήσεις της X, ο Ε.Μ.Π. της παραπάνω κατανομής **δεν είναι συνεπής**.




Πηγή: https://radfordneal.wordpress.com/2008/08/09/inconsistent-maximum-likelih

# Μέγιστη Πιθανοφάνεια $\not\rightarrow$ Μοναδικότητα

**Παράδειγμα 3**
Αν X ~ Laplace(μ, β) με $f(x) = \frac{1}{2\beta}\exp\left(-\frac{|x-\mu|}{\beta}\right)$, με άγνωστες παραμέτρους μ, β

και $x_1, x_2, \ldots, x_n$, ένα δείγμα τιμών, τότε αποδεικνύεται ότι ο εκτιμητής μέγιστης πιθανοφάνειας για την παράμετρο μ είναι η διάμεσος του δείγματος[1].

Καθώς η διάμεση τιμή δεν είναι μοναδική όταν το n είναι άρτιος, καταλαβαίνουμε ότι ο εκτιμητής δεν ορίζεται με μοναδικό τρόπο.

(1) Μία απόδειξη είναι διαθέσιμη εδώ: https://math.stackexchange.com/questions/240496/finding-the-maximum-likelihood-estimator

# Εκτιμητές μέγιστης πιθανοφάνειας

**Άσκηση 3.** Έστω X μία ΤΜ με κατανομή

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x|}{\beta}}$$

Επιλέγουμε δείγμα n τιμών από την κατανομή της X. Να βρεθεί ο Εκτιμητής Μέγιστης Πιθανοφάνειας της παραμέτρου β.

# Εκτιμητές μέγιστης πιθανοφάνειας

**Άσκηση 4.** Μία ΤΜ Χ, ακολουθεί την κατανομή Pareto με σ.π.π.

$$f(x) = \theta x^{-\theta-1},\ x \geq 1,\ \theta > 1.$$

Επιλέγουμε δείγμα n τιμών από την κατανομή της Χ. Να βρεθεί ο Εκτιμητής Μέγιστης Πιθανοφάνειας της παραμέτρου θ.

# Εκτιμητές μέγιστης πιθανοφάνειας

**Άσκηση 5.** Για μία ΤΜ Χ, γνωρίζουμε ότι Χ ~ U(0, θ). Επιλέγουμε δείγμα n τιμών από την κατανομή της Χ. Να βρεθεί ο Εκτιμητής Μέγιστης Πιθανοφάνειας της παραμέτρου θ.

# Εκτιμητές μέγιστης πιθανοφάνειας

**Εκτίμηση AR(1): MLE**

Μία χρονοσειρά που ακολουθεί ένα AR(1) μοντέλο $X_t = \delta + \phi X_{t-1} + w_t$ , $w_t \sim N(0, \sigma^2)$ αποτελείται από εξαρτημένες παρατηρήσεις.

Ωστόσο, τα σφάλματα $w_t$, είναι iid ακολουθία με σ.π.π. $f_w(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$.

Άρα, αν $W = (w_1, w_2, \ldots, w_n)$ η ακολουθία των σφαλμάτων, τότε η κοινή συνάρτηση πυκνότητας θα είναι το γινόμενο των επιμέρους.

$$L(W, \sigma) = f((w_1, w_2, \ldots, w_n \mid \sigma) = f(w_1 \mid \sigma)\, f(w_2 \mid \sigma) \ldots f(w_n \mid \sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i^2\right)$$

Όμως, $w_t = X_t - \delta - \phi X_{t-1}$ και η τελευταία σχέση ξαναγράφεται

$$L(X, \sigma) = f((x_2, x_3, \ldots, x_n) \mid x_0, \sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \left(x_i - \delta - \varphi x_{i-1}\right)^2\right)$$

Συμπεραίνουμε, ότι η συνάρτηση πιθανοφάνειας είναι

$$l(\sigma) = -\ln L(X, \sigma) = \frac{n}{2}\ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n} \left(x_i - \delta - \varphi x_{i-1}\right)^2 + c$$

# Εκτιμητές μέγιστης πιθανοφάνειας

**Εκτίμηση AR(1): MLE**

Η βελτιστοποίηση της $l(\sigma) = -\ln L(X, \sigma) = \frac{n}{2}\ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \delta - \varphi x_{i-1})^2 + c$

οδηγεί στη λύση:

$$\hat{\varphi} = \frac{\sum_{i=1}^{n}(x_{i-1} - \bar{x})(x_i - \bar{x}_+)}{\sum_{i=0}^{n-1}(x_i - \bar{x})^2}, \quad \hat{\delta} = \bar{x}_+ - \hat{\varphi}\,\bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\delta} - \hat{\varphi}\,x_{i-1})^2$$

όπου

$$\bar{x} = \frac{1}{n}\sum_{i=0}^{n-1} x_i, \quad \bar{x}_+ = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Σημείωση: Η παραπάνω διαδικασία γενικεύεται στην περίπτωση των AR(p) μοντέλων.

# Εκτιμητές μέγιστης πιθανοφάνειας

**Εναλλακτική Εκτίμηση AR(1): Exact MLE**

Από τη θεωρία πιθανοτήτων γνωρίζουμε ότι για θ = (δ, φ, $\sigma^2$) είναι:

$$f(x_0, x_1, x_2, \ldots, x_n|\theta) = f(x_0|\theta) f(x_1|x_0, \theta) \ldots f(x_n|x_{n-1}, x_{n-2}, \ldots, x_0, \theta) = f(x_0|\theta) \prod_{i=1}^{n} f(x_i|x_{i-1}, x_{i-2}, \ldots, x_0, \theta)$$

Καθώς η χρονοσειρά $X_t = \delta + \phi X_{t-1} + w_t$ είναι στάσιμη, είναι $X_0 \sim N(\mu, Var(X_t))$.

Όμως, $\mu = \delta / (1 - \phi)$, και $Var(X_t) = \sigma^2 / (1 - \phi^2)$, δηλαδή $X_0 \sim N(\delta / (1 - \phi), \sigma^2 / (1 - \phi^2))$ με

$$f(x_0|\theta) = \frac{\sqrt{1 - \varphi^2}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_0 - \delta/(1 - \varphi))^2}{2\sigma^2/(1 - \varphi^2)}\right)$$

Περαιτέρω, $x_n \mid x_{n-1}, x_{n-2}, \ldots, x_1 = x_n \mid x_{n-1} = N(\delta + \phi x_{n-1}, \sigma^2)$ και

$$f(x_n|x_{n-1}, x_{n-2}, \ldots, x_1, \sigma) = f(x_n|x_{n-1}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n - \delta - \varphi x_{n-1})^2}{2\sigma^2}\right)$$

# Εκτιμητές μέγιστης πιθανοφάνειας

**Εναλλακτική Εκτίμηση AR(1): Exact MLE**

Η συνάρτηση πιθανοφάνειας είναι η:

$$l(\sigma) = -\ln L(\sigma | X) = \frac{1}{2}\ln\frac{\sigma^2}{1-\varphi^2} + \frac{1}{2}n\ln\sigma^2 + \frac{\left(x_0 - \delta/(1-\varphi)\right)^2}{2\sigma^2\left(1-\varphi^2\right)} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \delta - \varphi x_{i-1}\right)^2 + c$$

Ο εντοπισμός των ακρότατων στην περίπτωση αυτή γίνεται με αριθμητικές – προσεγγιστικές μεθόδους.

# Σχέση μεταξύ MLE και Least Squares

Για το παράδειγμα AR(1), η συνάρτηση πιθανοφάνειας προέκυψε να είναι, μία από τις παρακάτω:

$$l(\sigma) = -\ln L(X, \sigma) = \frac{n}{2}\ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \delta - \varphi x_{i-1})^2 + c$$

$$l(\sigma) = -\ln L(\sigma | X) = \frac{1}{2}\ln\frac{\sigma^2}{1-\varphi^2} + \frac{1}{2}n\ln\sigma^2 + \frac{(x_0 - \delta/(1-\varphi))^2}{2\sigma^2(1-\varphi^2)} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \delta - \varphi x_{i-1})^2 + c$$

Παρατηρούμε πως οι συναρτήσεις αυτές αποκτούν ελάχιστη τιμή της όταν ελαχιστοποιείται το

$$\sum_{i=1}^{n}(x_i - \delta - \varphi x_{i-1})^2$$

Το τελευταίο άθροισμα το αναγνωρίζουμε ως το άθροισμα των τετραγωνικών σφαλμάτων το οποίο ελαχιστοποιείται στην μέθοδο των ελαχίστων τετραγώνων.

Συμπεραίνουμε πως, όταν $w_i \sim N(0, \sigma^2)$, τότε οι μέθοδοι μέγιστης πιθανοφάνειας και ελαχίστων τετραγώνων, οδηγούν στο ίδιο αποτέλεσμα.

# Πρόβλεψη

# Πρόβλεψη

Γνωρίζοντας τα στοιχεία $X_{1:n} = X_1, X_2, \ldots, X_n$, τι εκτίμηση μπορούμε να κάνουμε για την $X_{n+1}$;

Επιλέγουμε ως καλύτερη δυνατή πρόβλεψη Z, αυτή που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (MSE) $E((X_{n+1} - Z)^2)$.

Αποδεικνύεται ότι $\min[E((X_{n+1} - Z)^2)] = E(X_{n+1} \mid X_{1:n}) = X^*_{n+1}$.

Ως εκ τούτου η τιμή $\mathbf{X^*_{n+1} = E(X_{n+1} \mid X_{1:n})}$ θεωρείται η καλύτερη δυνατή εκτίμηση.

# Πρόβλεψη

**Παράδειγμα Ι**
Για την AR(1) χρονοσειρά $X_t = \delta + \phi X_{t-1} + w_t$, η εκτίμηση της n+1 τιμής από τις πρώτες n είναι η

$$\begin{aligned} X^*_{n+1} &= E(X_{n+1} \mid X_{1:n}) \\ &= E(\delta + \phi X_n + w_{n+1} \mid X_{1:n}) \\ &= \delta + \phi E(X_n \mid X_{1:n}) \\ &= \delta + \phi X_n \end{aligned}$$

Το σφάλμα της πρόβλεψης είναι το $w_{n+1}$ και η διακύμανσή του είναι $\sigma^2$.

Η εκτίμηση της n+2 τιμής από τις πρώτες n είναι

$$X^*_{n+2} = E(X_{n+2} \mid X_{1:n}) = E(\delta + \phi X_{n+1} + w_{n+2} \mid X_{1:n}) = \delta + \phi\, \mathbf{X^*_{n+1}} = \delta(1+ \phi) + \phi^2 X_n$$

Το σφάλμα της εκτίμησης είναι $w_{n+2} + \phi w_{n+1}$ με διακύμανση $(1 + \phi^2)\sigma^2$.

# Πρόβλεψη

**Παράδειγμα ΙΙ**
Για την MA(1) χρονοσειρά $X_t = \delta + \theta w_{t-1} + w_t$, η εκτίμηση της n+1 τιμής από τις πρώτες n είναι η

$$
\begin{aligned}
X^*_{n+1} &= E(X_{n+1} \mid X_{1:n}) \\
&= E(\delta + \theta w_n + w_{n+1} \mid X_{1:n}) \\
&= \delta + \theta E(w_n \mid X_{1:n}) + E(w_{n+1} \mid X_{1:n}) \\
&= \delta + \theta w_n
\end{aligned}
$$

Το σφάλμα της πρόβλεψης είναι το $w_{n+1}$ και η διακύμανσή του είναι $\sigma^2$.

# Αξιολόγηση των προβλέψεων

Μετά την ανάπτυξη ενός ARIMA μοντέλου, ακολουθεί ο διαγνωστικός έλεγχος.

Μεσο Σφάλμα $ME = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$

Ρίζα Μέσου Τετραγωνικού Σφάλματος $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2}$

Μεσο Απόλυτο Σφάλμα $MAE = \frac{1}{n}\sum_{i=1}^{n} |\varepsilon_i|$

Μεσο Ποσοστιαίο Σφάλμα $MPE = \frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_i}{x_i} 100$

Μεσο Απολυτο Ποσοστιαίο Σφάλμα $MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|\varepsilon_i|}{x_i} 100$