# Διακριτά Αντίστροφα Προβλήματα
## 2021-2022

Διάλεξη 7

Υπολογισμοί με μεθόδους βέλτιστης πιθανοφάνειας

Μιχάλης Ταρουδάκης

$\theta$
,

# Συνάρτηση μέγιστης πιθανοφάνειας

Εκτελούμε ένα πείραμα $N$ φορές και κάθε φορά παίρνουμε μέτρηση κάποιας ποσότητας $d_i$ που θα μας χρειαστεί στην επίλυση ενός αντίστροφου προβλήματος $d_i$

$P_d(d_i;\theta)$ Είναι η πιθανότητα να πάρει η μέτρηση $d_i$ *την τιμή* $\theta$

Η πιθανότητα να πάρουν όλες οι μετρήσεις την τιμή αυτή για ανεξάρτητες μεταξύ τους πιθανότητες είναι

$$P_d(d_1, d_2, \ldots . d_N;\theta) = P_d(d_1;\theta)\cdots P_d(d_N;\theta)$$

# Συνάρτηση μέγιστης πιθανοφάνειας

*συνάρτηση πιθανοφάνειας (likelihood function)*

$$L(\theta;\vec{d}) = P_d(d_1, d_2, ....d_N; \theta) = P_d(d_1; \theta) \cdots P_d(d_N; \theta)$$

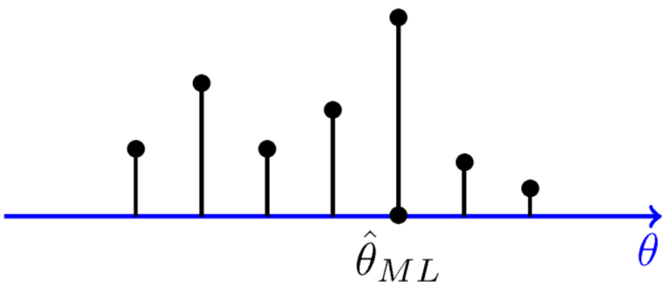
*εκτίμηση μέγιστης πιθανοφάνειας*; (*maximum likelihood estimate*) $\hat{\theta}$

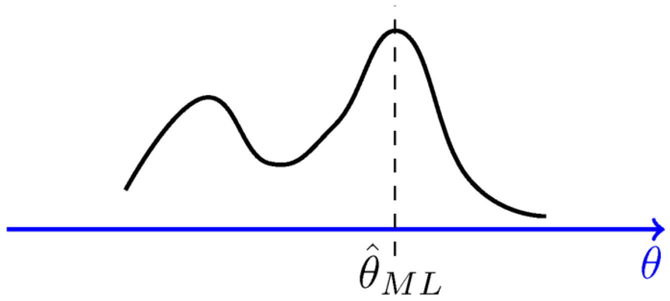
$$\hat{\theta}: \quad \frac{dL}{d\theta} = 0, \quad \frac{d^2 L}{d\theta^2} < 0$$

# Συνάρτηση μέγιστης πιθανοφάνειας

$$\hat{\theta}: \quad \frac{dL}{d\theta} = 0, \quad \frac{d^2L}{d\theta^2} < 0$$

$$L(\theta; d_1, d_2, \ldots, d_N)$$



# Συνάρτηση μέγιστης πιθανοφάνειας

Εάν $\vec{d}$ ακολουθεί κανονική κατανομή με μέση τιμή $m$ και διασπορά $\sigma$

$$\boxed{L(m;\sigma^2/\vec{d})} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(d_1-m)}{2\sigma^2})\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(d_2-m)}{2\sigma^2})\right)\cdots$$

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(d_N-m)}{2\sigma^2})\right) = \boxed{\frac{1}{\sqrt{(2\pi\sigma^2)^N}}\exp\left[-\frac{1}{2\sigma^2}\sum_{\iota=1}^{N}(d_i-m)^2\right]}$$

# Συνάρτηση μέγιστης πιθανοφάνειας

Εάν $\vec{d}$ ακολουθεί κανονική κατανομή με μέση τιμή $m$ και διασπορά $\sigma$

Λογαριθμική έκφραση συνάρτησης Μ.Π.

$$L' = \ln(L) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{\iota=1}^{N}(d_i - m)^2$$

$$\frac{\partial L'}{\partial m} = \frac{1}{\sigma^2}\sum_{\iota=1}^{N}(d_i - m) = \frac{1}{\sigma^2}\left(\sum_{\iota=1}^{N}d_i - Nm\right)$$

$$\frac{\partial L'}{\partial \sigma^2} = -\frac{N}{\sigma^2} + \frac{1}{\left(\sigma^2\right)^2}\sum_{\iota=1}^{N}(d_i - m)^2 = \frac{N}{\left(\sigma^2\right)^2}\left(\sigma^2 - \frac{1}{N}\sum_{\iota=1}^{N}(d_i - m)^2\right)$$

# Συνάρτηση μέγιστης πιθανοφάνειας

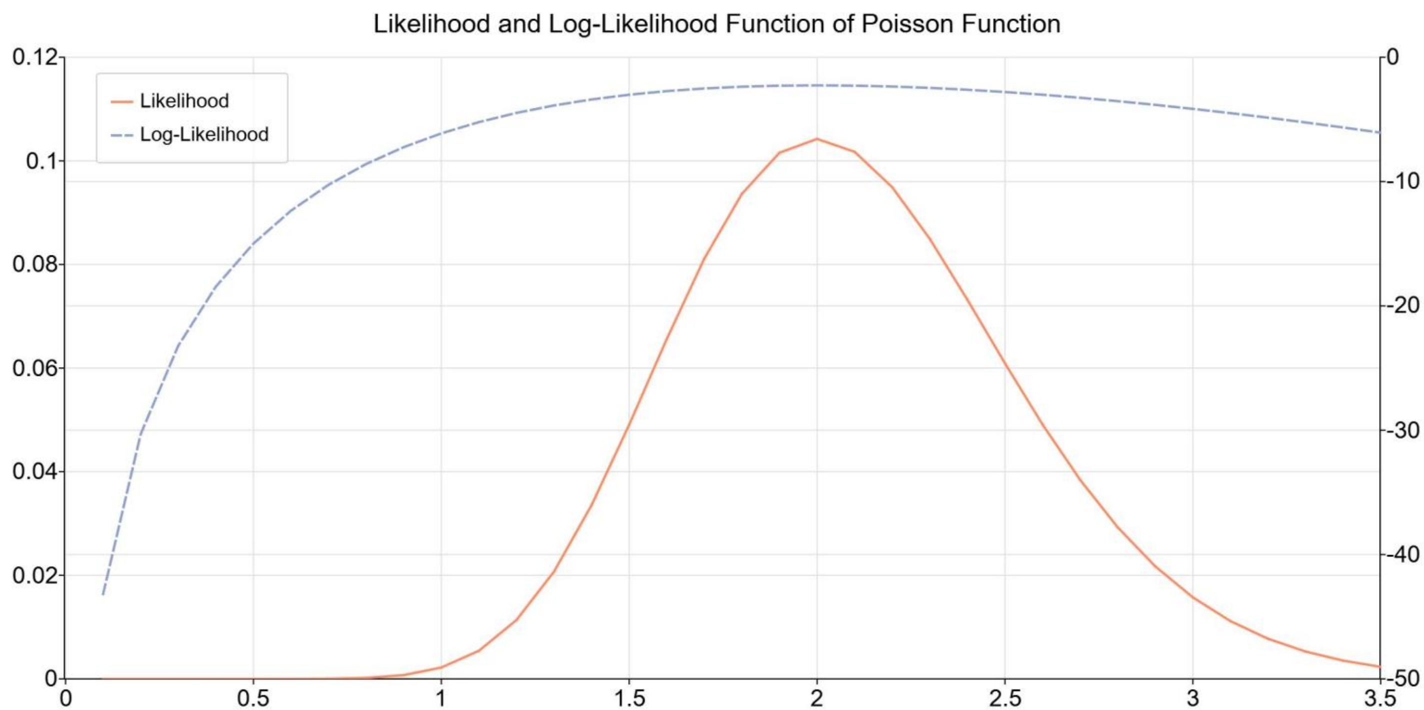
$$L' = \ln(L) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{\iota=1}^{N}(d_i - m)^2$$

Παραγωγίζοντας τη συνάρτηση μέγιστης πιθανοφάνειας ως προς τη μέση τιμή και την διακύμανση

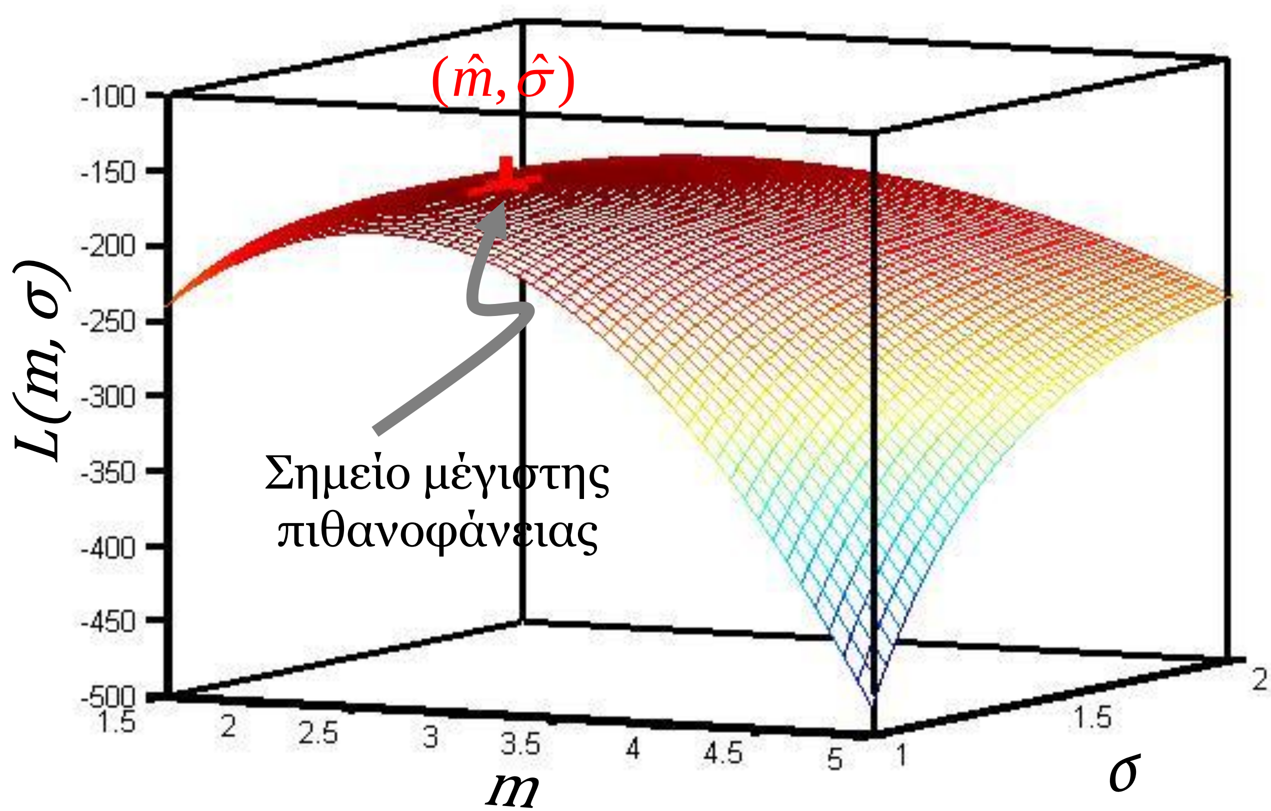
$$m^{est} = \frac{1}{N}\sum_{i=1}^{N} d_i \Longrightarrow \hat{m}$$

$$\left(\sigma^{est}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i - m^{est})^2 \Longrightarrow \hat{\sigma}^2$$

# Συνάρτηση μέγιστης πιθανοφάνειας



# Συνάρτηση μέγιστης πιθανοφάνειας



# Γραμμικό αντίστροφο πρόβλημα

$$\mathbf{Gm} = \mathbf{d}$$

Υποθέτουμε ότι τα δεδομένα μας (μετρήσεις) ακολουθούν την πολυπαραγοντική κανονική κατανομή

$$P(\mathbf{d}) \propto \exp\left[-\tfrac{1}{2}(\mathbf{d}-\mathbf{Gm})^T[\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{d}-\mathbf{Gm})\right]$$

$$\mathbf{d}-\mathbf{Gm} = \mathbf{d}^{obs} - \mathbf{d}^{pre} = \mathbf{e}$$

Θα θεωρήσουμε ότι $[\operatorname{cov}\mathbf{d}]$ είναι γνωστός

# Γραμμικό αντίστροφο πρόβλημα

$$\mathbf{Gm} = \mathbf{d}$$

Μεγιστοποίηση $P(\mathbf{d})$ : Ελαχιστοποίηση $(\mathbf{d} - \mathbf{Gm})^T [\operatorname{cov} \mathbf{d}]^{-1} (\mathbf{d} - \mathbf{Gm})$

Ζυγισμένο λάθος $E^w = \mathbf{e}^\mathbf{T} \mathbf{W_e} \mathbf{e} = \left(\mathbf{d} - \mathbf{Gm}\right)^T \mathbf{W_e} \left(\mathbf{d} - \mathbf{Gm}\right)$

Επομένως η λύση μέγιστης πιθανοφάνειας συμπίπτει με τη ζυγισμένη λύση ελαχίστων τετραγώνων

$$\mathbf{W}_e = [\operatorname{cov} \mathbf{d}]^{-1}$$

# Γραμμικό αντίστροφο πρόβλημα

$$\mathbf{Gm} = \mathbf{d}$$

Μεγιστοποίηση $P(\mathbf{d})$ : Ελαχιστοποίηση $(\mathbf{d}-\mathbf{Gm})^T[\mathrm{cov}\,\mathbf{d}]^{-1}(\mathbf{d}-\mathbf{Gm})$

Εάν $[\mathrm{cov}\,\mathbf{d}] = \sigma_d^2\mathbf{I}$

Λύση ελαχίστων τετραγώνων !

Εάν οι μετρήσεις είναι μεν ασυσχέτιστες αλλά έχουν διαφορετική διασπορά $\sigma_{di}$ το λάθος πρόβλεψης θα είναι

$$E^w = \sum_{i=1}^{N} \sigma_{di}^{-2} e_i^2$$

$$e_i = \left(d_i^{obs} - d_i^{pre}\right)$$

Οι μετρήσεις με μικρότερη αναμενόμενη διασπορά μετράνε περισσότερο

# Γραμμικό αντίστροφο πρόβλημα

$$\mathbf{Gm} = \mathbf{d}$$

Μεγιστοποίηση $P(\mathbf{d})$ : Ελαχιστοποίηση $(\mathbf{d} - \mathbf{Gm})^T [\operatorname{cov}\mathbf{d}]^{-1} (\mathbf{d} - \mathbf{Gm})$

**Λύση με βάση την αρχή μέγιστης πιθανοφάνειας :**

Εάν τα δεδομένα ακολουθούν γκαουσιανή στατιστική, λύστε το γραμμικό αντίστροφο πρόβλημα με τη χρήση της νόρμας ζυγισμένου λάθους, με τον πίνακα ζύγισης να είναι ο $[\operatorname{cov}\mathbf{d}]^{-1}$

Η προσέγγιση ελαχίστων τετραγώνων μας δίδει την λύση εάν τα δεδομένα μας είναι ασυσχέτιστα έχουν ίδια διακύμανση και ικανοποιούν την Γκαουσιανή στατιστική.

# Εκ προοιμίου κατανομές

Εάν το πρόβλημα είναι υπο-ορισμένο, δεν υφίσταται λύση ελαχίστων τετραγώνων

Θα μπορούσε να χρησιμοποιηθεί μια εκ προοιμίου (a-priori) πληροφορία που να δίδεται με τη μορφή συνάρτησης πιθανότητας

$$P_A(\mathbf{m})$$

Η πληροφορία αυτή μας δίδει την πιθανότητα η προς ανάκτηση τιμή της παραμέτρου να είναι κοντά σε μία μέση τιμή ενώ η διακύμανση μας υποδηλώνει την αβεβαιότητα ως προς την αξία της εκ προοιμίου πληροφορίας

# Εκ προοιμίου κατανομές

$P_A(\mathbf{m})$ Πληροφορία για τις προς εκτίμηση παραμέτρους

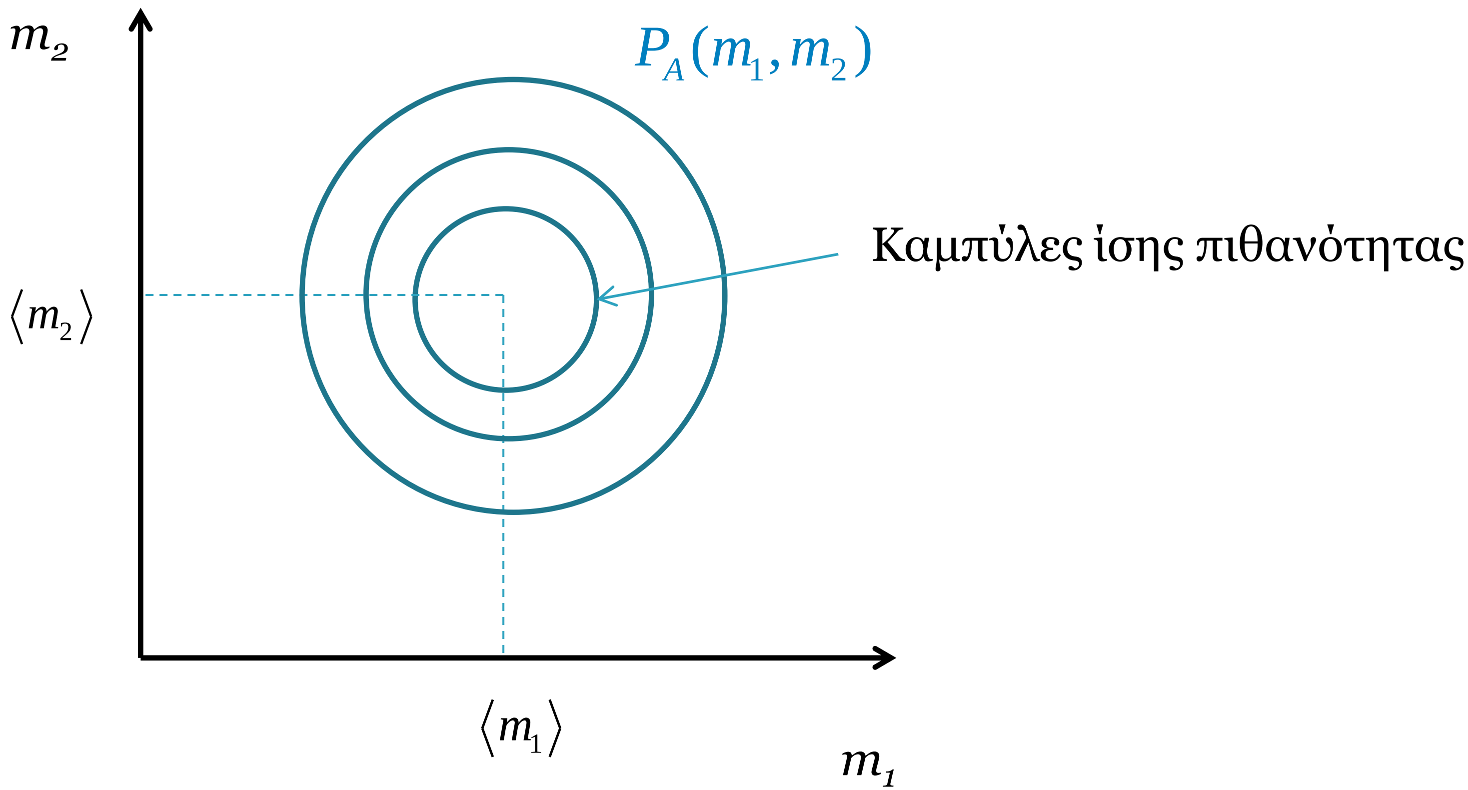
$P\left(\mathbf{d}_{obs}\right)$ Πληροφορία για την ποιότητα των μετρήσεων

Αξιοποίηση της εκ προοιμίου πληροφορίας για τις μετρήσεις σε συνδυασμό με την εκ προοιμίου εκτίμηση της ποιότητας των μετρήσεων μας οδηγεί σε μία από κοινού συνάρτηση πιθανότητας με ένα διακριτό μέγιστο.

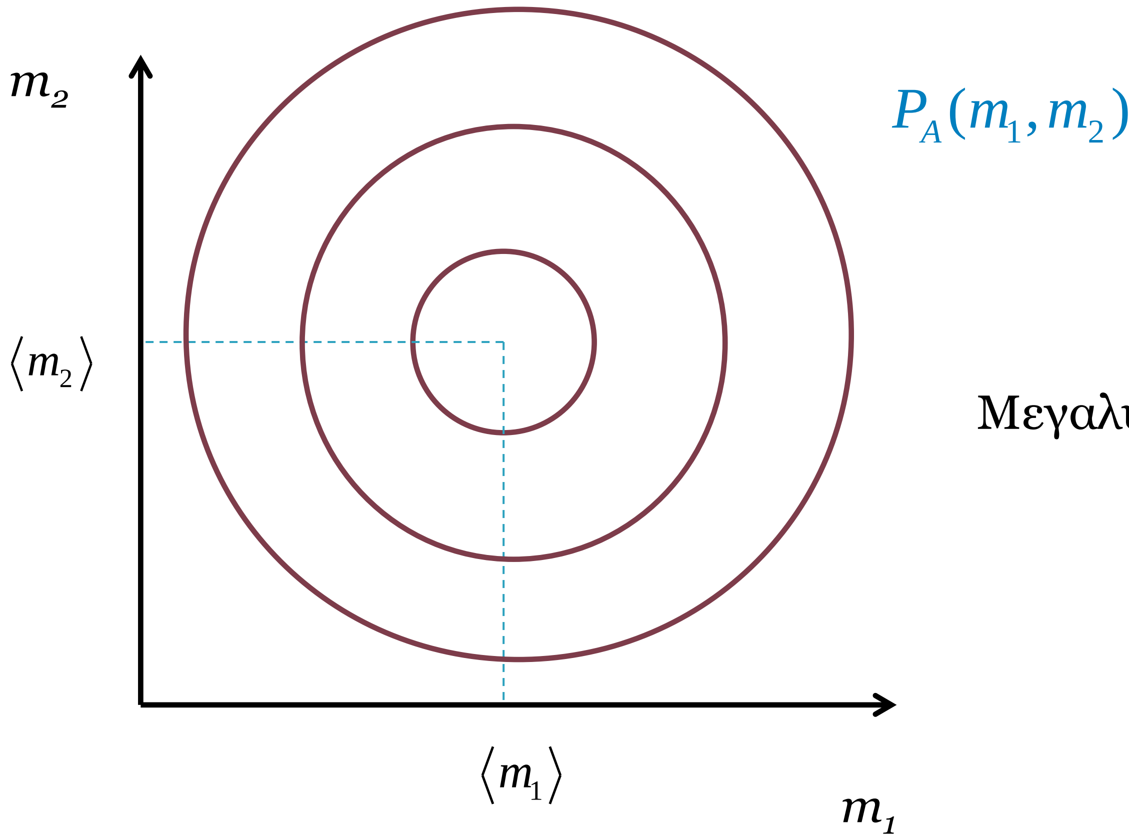
Οι εκ προοιμίου κατανομές μπορούν να πάρουν πολλές εναλλακτικές μορφές

Π.χ. μπορεί να θεωρηθεί ότι ψάχνομε να βρούμε παραμέτρους με πληροφορία ότι έχουν τιμές πολύ κοντά σε καποιο μέσο και μπορούν να περιγραφούν με κανονικές κατανομές γνωστής διακυμανσης

# Εκ προοιμίου κατανομές

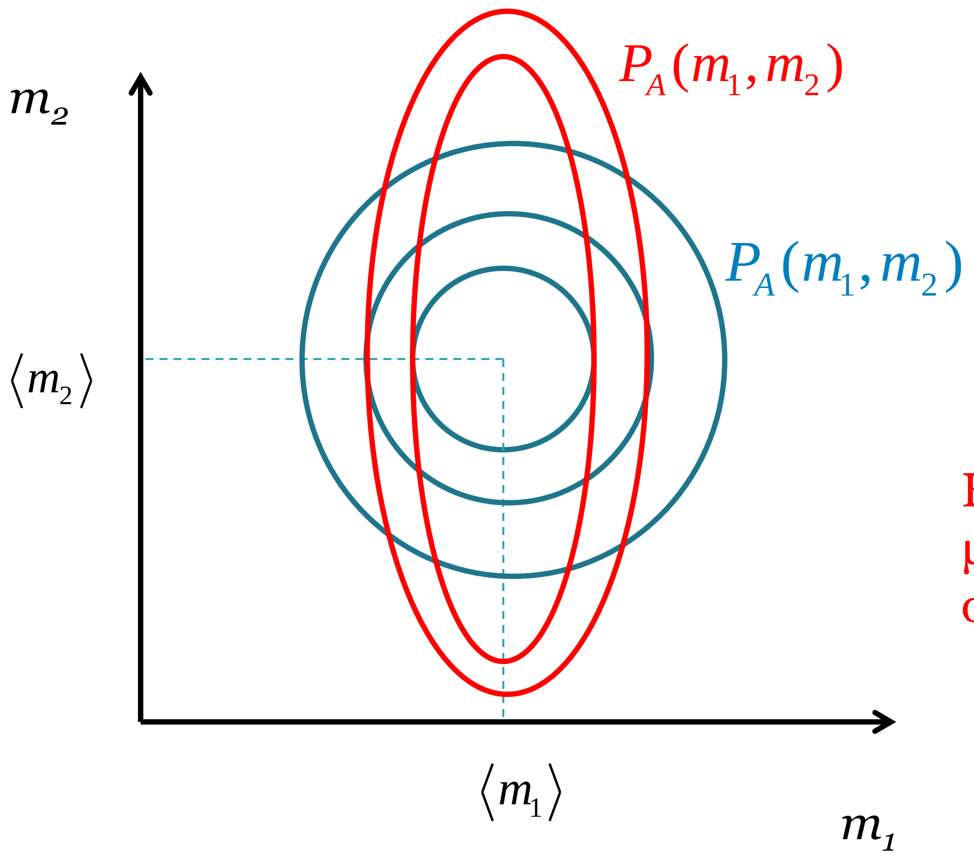


# Εκ προοιμίου κατανομές



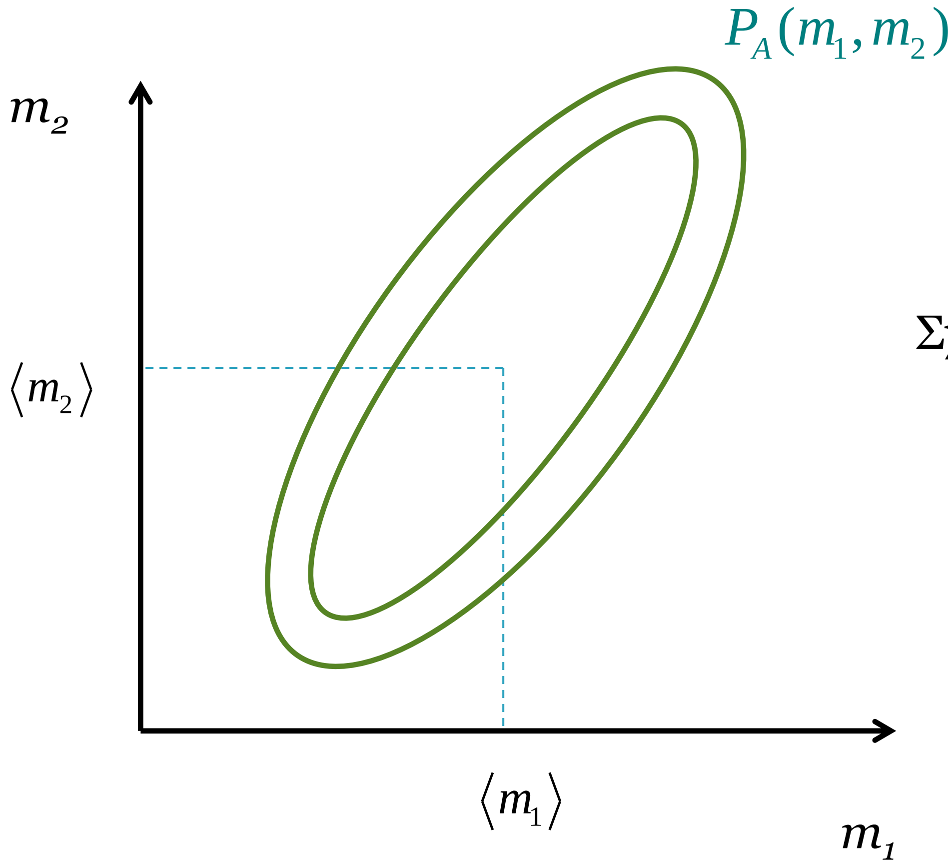

Μεγαλύτερη αβεβαιότητα

# Εκ προοιμίου κατανομές



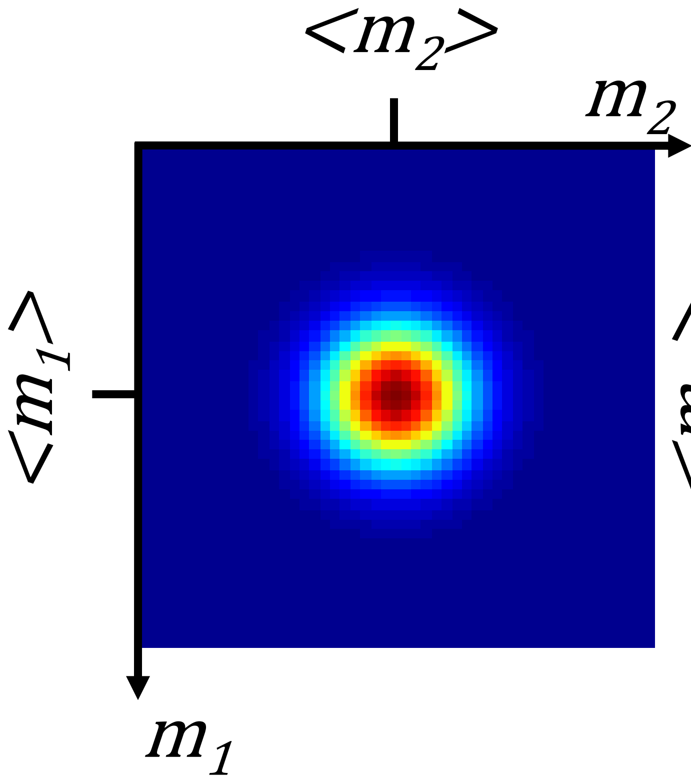

Η παράμετρος $m_2$ παρουσιάζει μεγαλύτερη αβεβαιότητα σε σχέση με την $m_1$

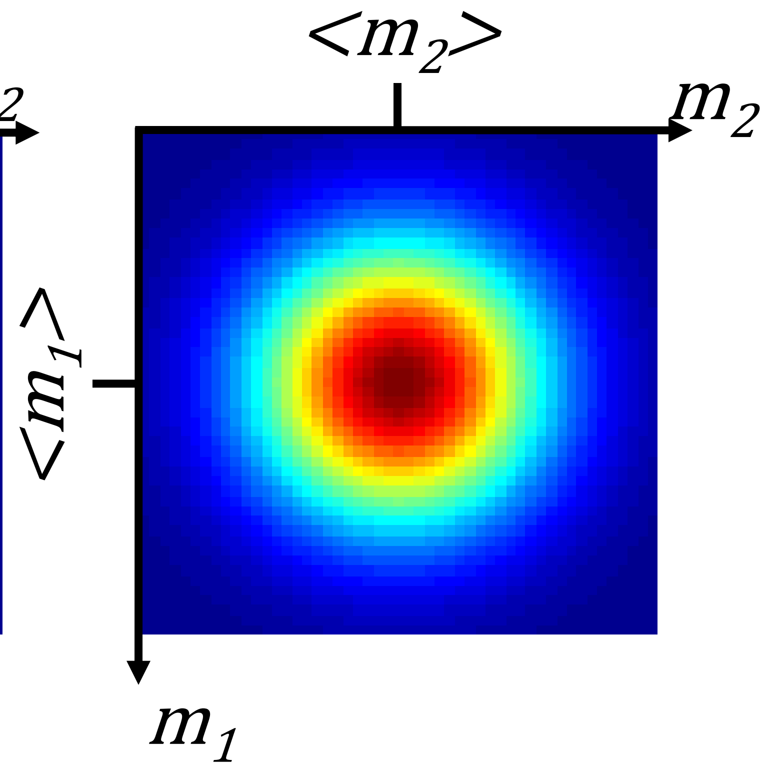
# Εκ προοιμίου κατανομές




Σχετιζόμενες παράμετροι

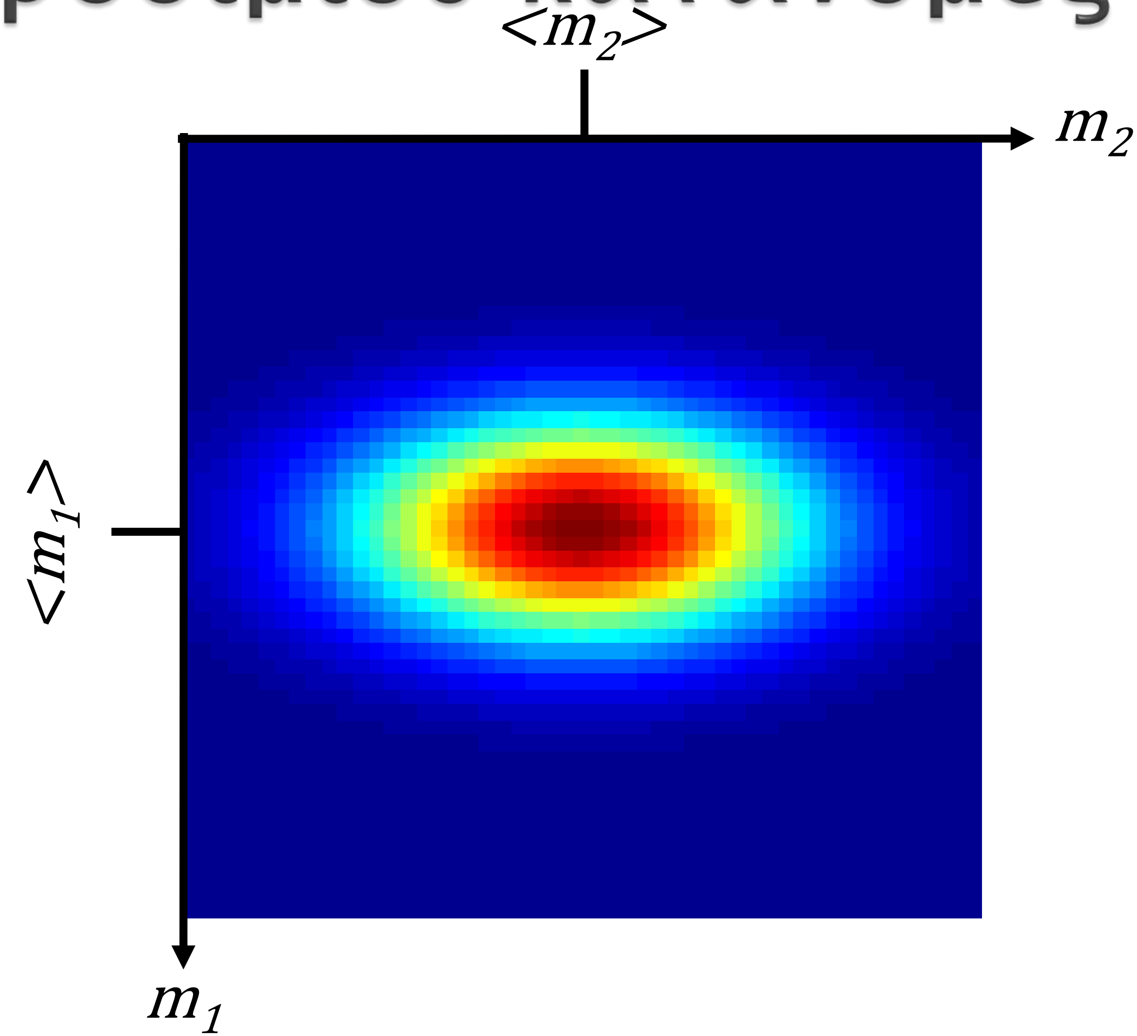
# Εκ προοιμίου κατανομές



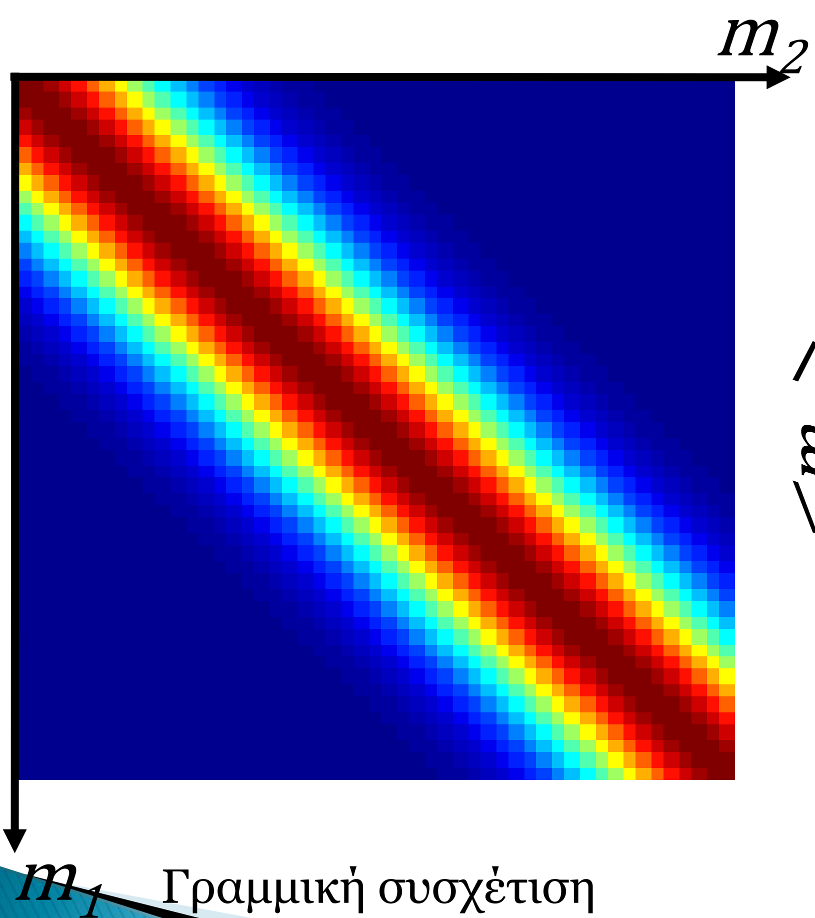




Μεγαλύτερη αβεβαιότητα

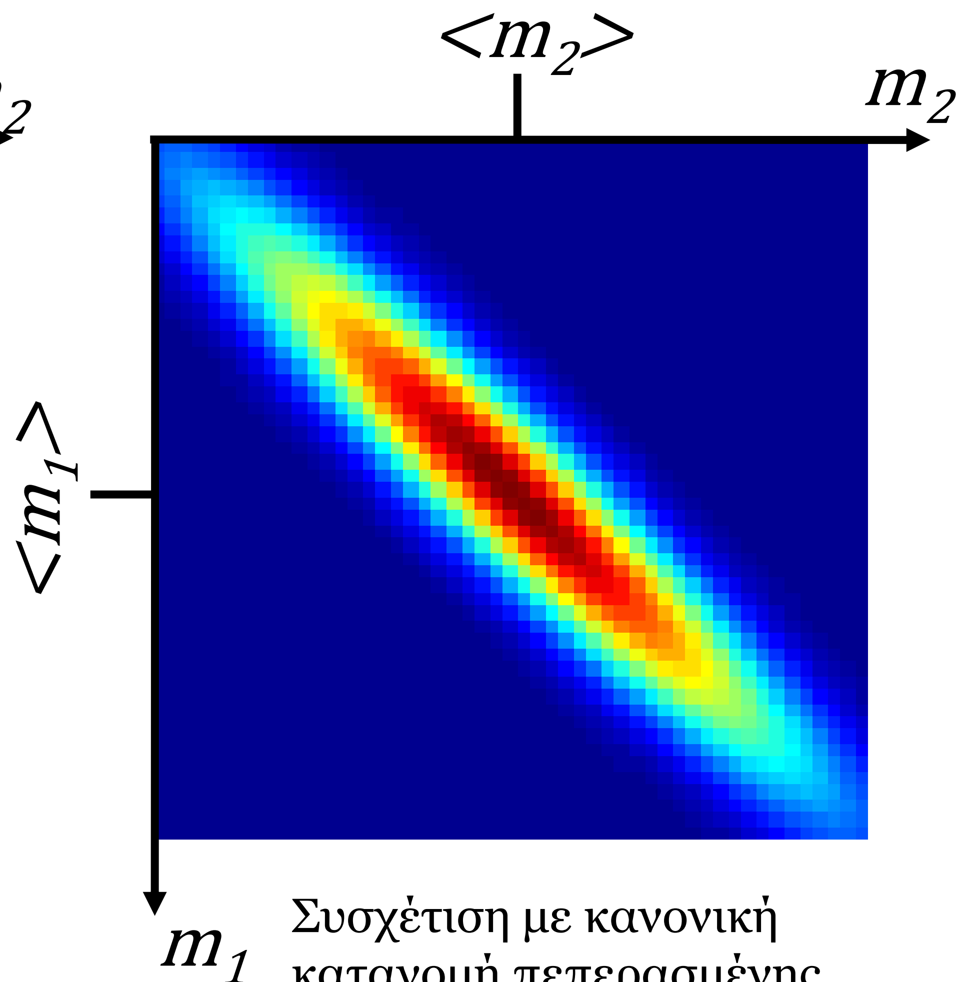
# Εκ προοιμίου κατανομές



# Εκ προοιμίου κατανομές




Γραμμική συσχέτιση




Συσχέτιση με κανονική κατανομή πεπερασμένης διακύμανσης

# Εκ προοιμίου κατανομές

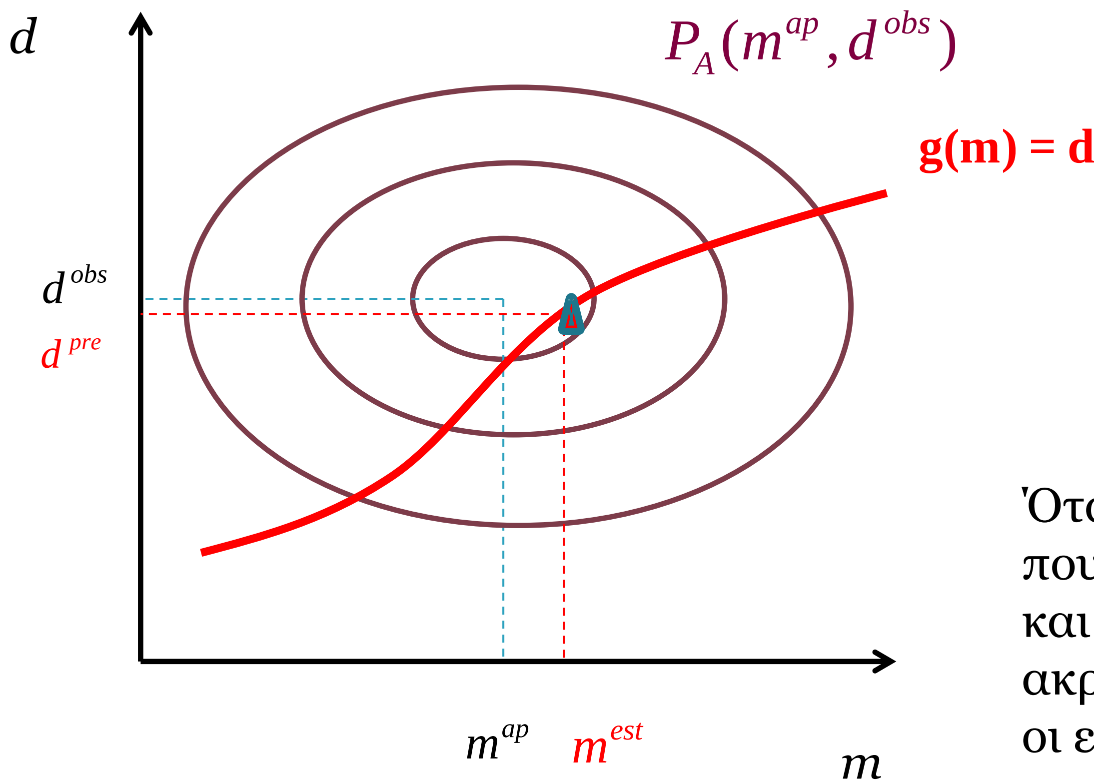
Εάν γνωρίζουμε τη συνάρτηση κατανομής πιθανότητας για δεδομένα και μετρήσεις και ότι αυτές είναι ασυσχέτιστες

$$P_A(\mathbf{m},\mathbf{d}) = P_A(\mathbf{m})P(\mathbf{d})$$

Μεγιστοποιώντας το $P_A$ παίρνομε τιμές που ουσιαστικά αναπλάθουν αυτή την πληροφορία (εκ προοιμίου πληροφορία). **Δεν συσχετίζουν όμως δεδομένα και παραμέτρους μέσω του μοντέλου.**

Εφαρμογή της αρχής μεγίστης πιθανοφάνειας απλά αναπλάθει τις εκ προοιμίου πληροφορίες !

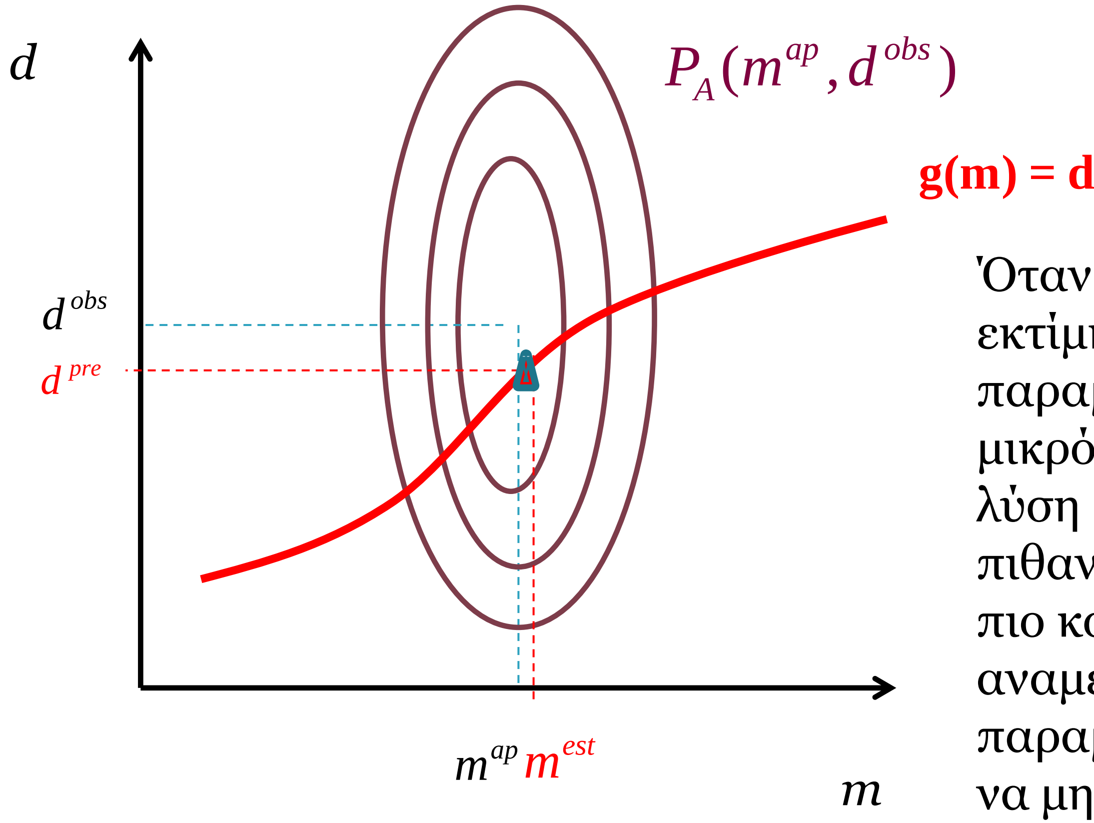
# Εκτίμηση μέγιστης πιθανοφάνειας για ακριβή θεωρία




Το μοντέλο (ακριβής θεωρία) μπορεί να μην είναι γραμμικό

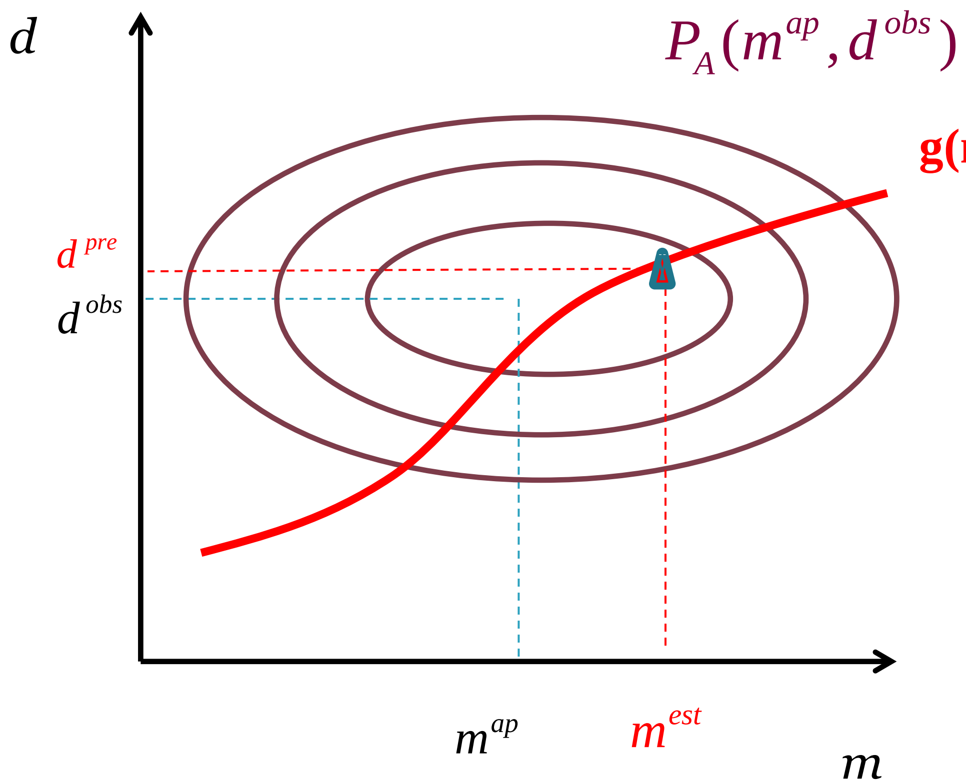
Όταν εφαρμοστεί η σχέση που συνδέει παραμέτρους και μετρήσεις μέσω ακριβούς θεωρίας, μπορεί οι εκτιμήσεις να διαφέρουν.

# Εκτίμηση μέγιστης πιθανοφάνειας για ακριβή θεωρία



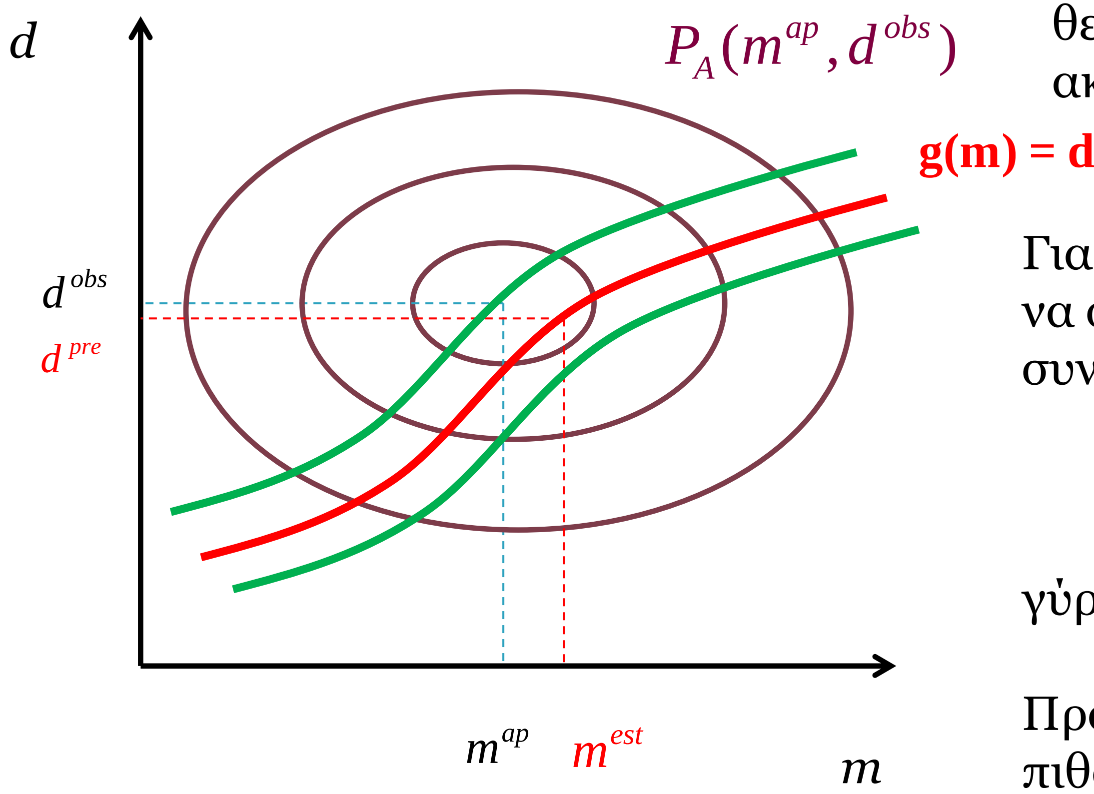

Όταν η εκ προοιμίου εκτίμηση για τις παραμέτρους έχει μικρότερη διακύμανση, η λύση μέγιστης πιθανοφάνειας θα βρεθεί πιο κοντά στην μέση αναμενόμενη τιμή της παραμέτρου αλλά μπορεί να μην αντιπροσωπεύει σωστά τα δεδομένα

# Εκτίμηση μέγιστης πιθανοφάνειας για ακριβή θεωρία




Όταν η εκ προοιμίου εκτίμηση για τις παραμέτρους έχει μεγαλύτερη διακύμανση, η λύση μέγιστης πιθανοφάνειας θα βρεθεί πιο κοντά στα δεδομένα αλλά μπορεί να είναι μακρυά από τις μέσες τιμές των παραμέτρων που προέρχονται από την εκ προοιμίου πληροφορία

# Εκτίμηση μέγιστης πιθανοφάνειας για μη ακριβή θεωρία




Υπάρχει πιθανότητα η θεωρία να μην είναι ακριβής

Για μη ακριβή θεωρία μπορεί να οριστεί μια αντίστοιχη συνάρτηση πιθανότητας

$$P_g(\mathbf{m}|\mathbf{d})$$

γύρω από την **g(m) = d**

Πρόκειται για υπό συνθήκη πιθανότητα.

# Εκτίμηση μέγιστης πιθανοφάνειας για μη ακριβή θεωρία

$P_g(\mathbf{m}|\mathbf{d})$ Δεδομένου ότι η υπό συνθήκη πιθανότητα είναι ασυσχέτιστη με την από κοινού συνάρτηση πιθανότητας παραμέτρων και μετρήσεων

$$P_T(\mathbf{m},\mathbf{d}) = P_g(\mathbf{m}|\mathbf{d})P_A(\mathbf{m},\mathbf{d})$$ «Ολική πιθανότητα»

Η διαφοροποίηση από την αρχή ελαχίστων τετραγώνων, εάν π.χ. αυτή εφαρμοστεί στη λύση του αντίστροφου προβλήματος, είναι ότι τώρα γίνεται ταυτόχρονα ανάκτηση των πλέον πιθανών τιμών παραμέτρων και μετρήσεων. Οι λύσεις μπορεί να μην είναι ίδιες

# Εκτίμηση μέγιστης πιθανοφάνειας για μη ακριβή θεωρία

$P_g(\mathbf{m}|\mathbf{d})$ Δεδομένου ότι η υπό συνθήκη πιθανότητα είναι ασυσχέτιστη με την από κοινού συνάρτηση πιθανότητας παραμέτρων και μετρήσεων

$$P_T(\mathbf{m},\mathbf{d}) = P_g(\mathbf{m}|\mathbf{d})P_A(\mathbf{m},\mathbf{d})$$

«Ολική πιθανότητα»

Εύρεση του σημείου μέγιστης πιθανοφάνειας ως προς τις παραμέτρους και μόνο

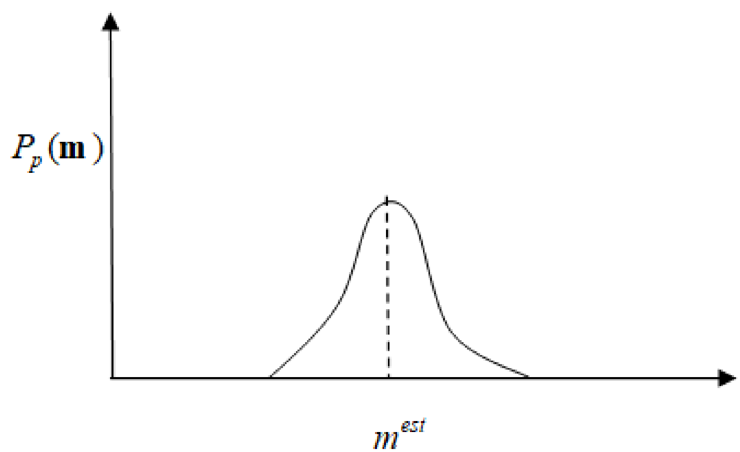
Άθροισμα της ολικής πιθανότητας κατά μήκος των γραμμών ίσης τιμής της παραμέτρου

$$P_p(\mathbf{m}) = \int P_T(\mathbf{m},\mathbf{d})\partial\mathbf{d}$$

# Εκτίμηση μέγιστης πιθανοφάνειας για μη ακριβή θεωρία

$$P_p(\mathbf{m}) = \int P_T(\mathbf{m}, \mathbf{d}) \partial \mathbf{d}$$

Εύρεση του σημείου μέγιστης πιθανοφάνειας ως προς τις παραμέτρους και μόνο : προβολή στο επίπεδο **d**=0




Εάν η θεωρία είναι «σχεδόν» ακριβής, παίρνομε πρακτικά τη λύση μέγιστης πιθανοφάνειας για ακριβή θεωρία

# Παραδείγματα

Θεωρούμε ότι δεν υπάρχει κάποια πληροφορία για την $P_A(\mathbf{m})$ και ότι τα δεδομένα ακολουθούν κανονική κατανομή.

$$P_A(\mathbf{d}) \propto \exp[-\tfrac{1}{2}(\mathbf{d}-\mathbf{d}^{obs})^T[\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{d}-\mathbf{d}^{obs})]$$

Ακριβές μοντέλο $\qquad P_g(\mathbf{m}|\mathbf{d}) = \delta[\mathbf{Gm} - \mathbf{d}]$

$$P_T(\mathbf{m},\mathbf{d}) \propto \exp[-\tfrac{1}{2}(\mathbf{d}-\mathbf{d}^{obs})^T[\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{d}-\mathbf{d}^{obs})]\delta[\mathbf{Gm} - \mathbf{d}]$$

$$P_T(\mathbf{m}) \propto \exp[-\tfrac{1}{2}(\mathbf{Gm}-\mathbf{d}^{obs})^T[\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{Gm}-\mathbf{d}^{obs})]$$

Η αρχή μέγιστης πιθανοφάνειας οδηγεί σε λύση ελαχίστων τετραγώνων

**Ακριβής γραμμική θεωρία με κανονική κατανομή δεδομένων**

# Παραδείγματα

$$P_A(\mathbf{m}) \propto \exp[-\tfrac{1}{2}(\mathbf{m}-<\mathbf{m}>)^T[\operatorname{cov}\mathbf{m}]^{-1}(\mathbf{m}-<\mathbf{m}>)]$$

$$P_A(\mathbf{d}) \propto \exp[-\tfrac{1}{2}(\mathbf{d}-\mathbf{d}^{obs})^T[\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{d}-\mathbf{d}^{obs})]$$

$$P_g(\mathbf{m}|\mathbf{d}) \propto \exp[-\tfrac{1}{2}(\mathbf{d}-\mathbf{Gm})^T[\operatorname{cov}\mathbf{g}]^{-1}(\mathbf{d}-\mathbf{Gm})]$$

$$P_T(\mathbf{m}) = P_g(\mathbf{m}|\mathbf{d})P_A(\mathbf{m})P_A(\mathbf{d})$$

Μπορεί να αποδειχθεί ότι η συνδυασμένη (τελική) πιθανότητα είναι επίσης κανονική

**Γενική γραμμική θεωρία με κανονική κατανομή δεδομένων και παραμέτρων**

# Παραδείγματα

$$\mathbf{m}^{est} = \langle \mathbf{m} \rangle + \mathbf{G}^{-g}[\mathbf{d}^{obs} - \mathbf{G}\langle \mathbf{m} \rangle] = \mathbf{G}^{-g}\mathbf{d} + [\mathbf{I}-\mathbf{R}]\langle \mathbf{m} \rangle$$

$$\mathbf{G}^{-g} = [\operatorname{cov}\mathbf{m}]\mathbf{G}^T\left\{[\operatorname{cov}\mathbf{d}] + [\operatorname{cov}\mathbf{g}] + \mathbf{G}[\operatorname{cov}\mathbf{m}]\mathbf{G}^T\right\}^{-1}$$

$$= \left\{\mathbf{G}^T[[\operatorname{cov}\mathbf{d}] + [\operatorname{cov}\mathbf{g}]]^{-1}\mathbf{G} + [\operatorname{cov}\mathbf{m}]^{-1}\right\}^{-1}\mathbf{G}^T\left\{[\operatorname{cov}\mathbf{d}] + [\operatorname{cov}\mathbf{g}]\right\}^{-1}$$

Οι εκτιμήσεις των παραμέτρων είναι γραμμικοί συνδυασμοί μετρήσεων και εκ προοιμίου τιμών των παραμέτρων

$$[\operatorname{cov}\mathbf{m}^{est}] = \mathbf{G}^{-g}[\operatorname{cov}\mathbf{d}]\mathbf{G}^{-gT} + [\mathbf{I}-\mathbf{R}][\operatorname{cov}\mathbf{m}][\mathbf{I}-\mathbf{R}]^T$$

Επί πλέον όρος

**Γενική γραμμική θεωρία με κανονική κατανομή δεδομένων και παραμέτρων**

# Ειδικές Περιπτώσεις

Ακριβής θεωρία και ακριβή δεδομένα $\sigma_d^2 = \sigma_g^2 = 0$

$$\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d}^{obs} + [\mathbf{I} - \mathbf{R}]\langle \mathbf{m} \rangle$$

$\mathbf{G}^{-g}$ Εξαρτάται από το είδος του προβλήματος

$\mathbf{G}^{-g} = \left[\mathbf{G}^T\mathbf{G}\right]^{-1}\mathbf{G}^T$ Υπερ-ορισμένο πρόβλημα

$\mathbf{G}^{-g} = \mathbf{G}^T\left[\mathbf{G}\mathbf{G}^T\right]^{-1}$ Υπο-ορισμένο πρόβλημα

Υπερορισμένο πρόβλημα : $[\mathbf{I} - \mathbf{R}]\langle \mathbf{m} \rangle = 0$

Η εκ προοιμίου πληροφορία σε ένα πρόβλημα που διέπεται από ακριβή θεωρία και ακριβή δεδομένα, επηρεάζει μόνο το υποορισμένο τμήμα του προβλήματος

# Ειδικές Περιπτώσεις

Μη ακριβής θεωρία και μη ακριβή δεδομένα (ακραία μη ρεαλιστική περίπτωση)

$$\sigma_d^2 \to \infty, \sigma_g^2 \to \infty$$

$$\mathbf{m}^{est} = \langle \mathbf{m} \rangle$$

# Ειδικές Περιπτώσεις

Μη ακριβής πληροφορία για τις παραμέτρους

$$\sigma_m^2 \to \infty$$

$$\mathbf{m}^{est} = \mathbf{G}^T[\mathbf{G}\mathbf{G}^T]^{-1}\mathbf{d}^{obs} + \{\mathbf{I} - \mathbf{G}^T[\mathbf{G}\mathbf{G}^T]^{-1}\mathbf{G}\}\langle\mathbf{m}\rangle$$

Η παραγόμενη λύση είναι ίδια με εκείνη που παίρνομε με πεπερασμένη πληροφορία για τις παραμέτρους αλλά ακριβή πληροφορία για τα δεδομένα και τη θεωρία

# Ανακεφαλαίωση

Αντιμετώπιση του γραμμικού αντίστροφου προβλήματος

Α: Ελαχιστοποίηση λάθους (ζυγισμένου)

$$\Phi' = \mathbf{e}^T \mathbf{W}_e \mathbf{e} + \varepsilon^2 \left[ \mathbf{m} - \langle \mathbf{m} \rangle \right]^T \mathbf{W}_m \left[ \mathbf{m} - \langle \mathbf{m} \rangle \right]$$

Β: Βελτιστοποίηση ποιότητας μετρήσεων και παραμέτρων (ζυγισμένης): Ελαχιστοποίηση της συνάρτησης

$$\Phi'' = \alpha_1 \mathrm{spread}(\mathbf{R}) + \alpha_2 \mathrm{spread}(\mathbf{N}) + \alpha_3 \mathrm{size}\left[ \mathrm{cov}_u\, \mathbf{m} \right]$$

Γ: Αξιοποίηση πιθανοθεωρητικών δεδομένων μετρήσεων, παραμέτρων και μοντέλου. Μεγιστοποίηση της συνάρτησης πυκνότητας πιθανότητας

$$P_T(\mathbf{m}, \mathbf{d})$$