Περιγραφική Στατιστική

Γιώργος Αραμπατζής
Τμήμα Μαθηματικών και Εφαρμοσμένων Μαθηματικών
Πανεπιστήμιο Κρήτης

Εαρινό 2024

Περίληψη

Οι σημειώσεις αποτελούν συμπληρωματικό υλικό στο μάθημα Περιγραφική Στατιστική όπως διδάχθηκε στο εαρινό εξάμηνο του 2024.

1 Εβδομάδα 01/13

Η Στατιστική είναι ο κλάδος των μαθηματικών που ασχολείται με την συλλογή, ανάλυση, παρουσίαση, και την ερμηνεία δεδομένων και την λήψη αποφάσεων. Χωρίζεται στις κατηγορίες περιγραφική (descriptive) και επαγωγική (inferential) στατιστική. Η περιγραφική στατιστική χρησιμοποιεί μεθόδους για την παρουσίαση και την περιγραφή των δεδομένων. Η επαγωγική στατιστική ασχολείται με την χρήση δειγμάτων με σκοπό την λήψη αποφάσεων ή την δυνατότητα να κάνουμε προβλέψεις.

Ένα συγκεκριμένο υποκείμενο ή αντικείμενο για το οποίο συλλέγουμε πληροφορίες ονομάζεται στοιχείο. Ένα χαρακτηριστικό του στοιχείου υπό μελέτη ονομάζεται μεταβλητή. Κάθε μεταβλητή είναι δυνατό να παίρνει πάνω από μία τιμές. Μέτρηση ή παρατήρηση είναι η τιμή μιας μεταβλητής. Η συλλογή στοιχείων μαζί με τις μεταβλητές και τις τιμές τους ονομάζεται σύνολο δεδομένων (dataset).

Το σύνολο όλων των στοιχείων μιας μελέτης ονομάζεται πληθυσμός (population) και η συλλογή της πληροφορίας γίνεται με απογραφή (census). Ένα υποσύνολο του πληθυσμού ονομάζεται δείγμα (sample) και η συλλογή της πληροφορίας γίνεται με έρευνα (survey).

Τις μεταβλητές τις συμβολίζουμε με κεφαλαία γράμματα (X, Y, Z), και τις τιμές τους με μικρά (x, y, z). Θεωρούμε ότι οι τιμές των μεταβλητών είναι τυχαίες.

1.1 Σύνολο δεδομένων

Στον Πίνακα 1 παρουσιάζεται ένα *σύνολο δεδομένων* (dataset). Η πρώτη γραμμή του πίνακα ονομάζεται επικεφαλίδα (header) και μας δίνει τις βασικές πληροφορίες του συνόλου δεδομένων. Η πρώτη στήλη αντιστοιχεί στα *στοιχεία* του συνόλου. Τα στοιχεία του συνόλου στον Πίνακα 1 είναι τα ονόματα των ανθρώπων που συμμετείχαν στη μελέτη. Το σύνολο έχει δύο *μεταβλητές* με ονόματα Height και Weight. Σε κάθε στοιχείο αντιστοιχεί μία *μέτρηση* για κάθε μεταβλητή.

Name	Height	Weight
George	100	180
Maria	60	160
John	80	190
⋮	⋮	⋮

Πίνακας 1: Τα πρώτα στοιχεία ενός συνόλου δεδομένων.

1.2 Τυχαίες Μεταβλητές

Δειγματικός χώρος είναι το σύνολο όλων των δυνατών αποτελεσμάτων ενός πειράματος, όλες οι τιμές που μπορεί να πάρει μια μεταβλητή. Συνήθως συμβολίζουμε τον δειγματικό χώρο με το γράμμα Ω. Ένα υποσύνολο του δειγματικού χώρο το ονομάζουμε ενδεχόμενο.

Παράδειγμα 1.1. *Θεωρήστε το πείραμα όπου ρίχνουμε ένα νόμισμα δύο φορές. Αν το νόμισμα έρθει κορώνα γράφουμε* K*, αλλιώς γράφουμε* Γ*. Ο δειγματικός χώρος είναι το σύνολο* $\Omega = \{KK, K\Gamma, \Gamma K, \Gamma\Gamma\}$.

Παράδειγμα 1.2. *Ρίχνουμε ένα ζάρι μέχρι να έρθει 6,*

$$\Omega = \{6, (1,6), (2,6) \ldots, (5,6), (1,1,6), (1,2,6), \ldots, (5,5,6), \ldots\}.$$

Παρατηρήστε ότι $|\Omega| = \infty$.

Ερώτηση 1.1. *Πόσο είναι το* $|\Omega|$ *στο Παράδειγμα 1.2 εάν κάνουμε το πολύ 5 ρίψεις;*

Ορισμός 1.1 (τυχαία μεταβλητή)**.** Τυχαία μεταβλητή X *ονομάζουμε μια συνάρτηση από το* Ω *στο* $\mathbb{R}$, $X : \Omega \to \mathbb{R}$*. Κάθε μεταβλητή ενός συνόλου δεδομένων μπορεί να συσχετιστεί με μία τυχαία μεταβλητή.*

Παράδειγμα 1.3. *Ρίχνουμε ένα ζάρι 3 φορές και καταγράφουμε το άθροισμα. Τότε* $\Omega = \{(i,j,k) : i,j,k = 1,\ldots,6\}$ *και* $X(\omega) = X(i,j,k) = i + j + k$.

Παράδειγμα 1.4. *Ρίχνουμε ένα κέρμα 3 φορές και μετράμε των αριθμό των κεφαλιών. Τότε*

$$\Omega = \{KKK, KK\Gamma, K\Gamma K, K\Gamma\Gamma, \Gamma KK, \Gamma K\Gamma, \Gamma\Gamma K, \Gamma\Gamma\Gamma\}$$

και $X(\omega) =$ *πλήθος των* K *στο* ω *για κάθε* $\omega \in \Omega$.

Παράδειγμα 1.5. *Ο χρόνος αναμονής του πελάτη σε ένα τηλεφωνικό κέντρο:* $\Omega = [0, \infty]$ *και* $X(\omega) = \omega$.

Ορισμός 1.2 (κατηγορίες μεταβλητών)**.** *Όταν μια μεταβλητή ενός συνόλου δεδομένων εκφράζεται αριθμητικά σύμφωνα με μία μονάδα μέτρησης, τότε ονομάζεται ποσοτική. Όταν περιγράφει χαρακτηριστικά του πληθυσμού που μεταβάλλονται κατά ποιότητα ή είδος, αλλά όχι κατά μέγεθος, τότε ονομάζεται ποιοτική. Οι ποσοτικές τυχαίες μεταβλητές χωρίζονται σε διακριτές και συνεχείς, αναλόγως αν το πεδίο τιμών της τυχαίας μεταβλητής είναι αριθμήσιμο ή όχι. Οι ποιοτικές μεταβλητές χωρίζονται σε διατάξιμες και ονομαστικές αναλόγως αν μπορεί να ισχύει μια σχέση διάταξης ανάμεσα τους ή όχι.*

Υπάρχουν δύο τρόποι για να εκφράσουμε μια ποιοτική μεταβλητή ως τυχαία μεταβλητή: integer encoding, και one-hot encoding. Στον integer encoding αντιστοιχούμε σε κάθε τιμής της μεταβλητής έναν ακέραιο.

Παράδειγμα 1.6. *Έστω μια μεταβλητή με τρεις τιμές, χαμηλή, μέση, και υψηλή θερμοκρασία. Μπορούμε σε κάθε τιμή να αντιστοιχήσουμε τους αριθμούς, 0, 1, και 2, αντίστοιχα.*

Στον one-hot encoding τρόπο βάζουμε όλες τις τιμές της μεταβλητής σε τυχαία σειρά. Σε κάθε τιμή αντιστοιχούμε ένα διάνυσμα με μέγεθος ίσο με το πλήθος των τιμών της μεταβλητής, που έχει μηδενικά σε κάθε θέση και 1 στην θέση που αντιστοιχεί στη σειρά της μεταβλητής στην διάταξη που ορίσαμε αρχικά.

Παράδειγμα 1.7. *Έστω μια μεταβλητή με τρεις τιμές, ήλιος, βροχή, και σύννεφα. Σε κάθε μία, με την σειρά που είναι γραμμένες στην προηγούμενη πρόταση, αντιστοιχούμε το διάνυσμα* $(1,0,0)$, $(0,1,0)$, *και* $(0,0,1)$.

Ερώτηση 1.2. *Τι πλεονεκτήματα έχει ο κάθε τρόπος κωδικοποίησης;*

Ορισμός 1.3 (σύνολα δεδομένων)**.** *Τα σύνολα δεδομένων χωρίζονται σε δύο κατηγορίες σύμφωνα με τον χρόνο συλλογής: α)* διαστρωματικά *(cross-sectional data), όταν τα δεδομένα συλλέγονται σε μια συγκεκριμένη χρονική περίοδο β)* χρονολογικά *(time-series data), όταν τα δεδομένα συλλέγονται σε μια μεγάλη χρονική περίοδο και ο χρόνος περιέχεται στα δεδομένα.*

1.3 Οργάνωση ποσοτικών δεδομένων

Στον Πίνακα 2 θεωρούμε την μεταβλητή που δίνει την ηλικία κάθε στοιχείου. Οργανώνουμε τα δεδομένα σε κλάσεις με σκοπό να γίνει ευκολότερη η μελέτη τους. Τις κλάσεις μπορούμε να τις ορίσουμε αυθαίρετα. Έναν τρόπο επιλογής θα περιγράψουμε παρακάτω.

$(37, M)$	$(24, F)$	$(22, F)$	$(20, M)$
$(18, M)$	$(22, M)$	$(21, F)$	$(18, F)$
$(19, F)$	$(19, F)$	$(34, F)$	$(33, F)$
$(22, F)$	$(28, M)$	$(19, M)$	$(19, F)$
$(30, M)$	$(20, F)$	$(22, M)$	$(34, M)$

Πίνακας 2: Οι τιμές για δύο μεταβλητές Age και Sex ενός συνόλου δεδομένων.

Αρχικά επιλέγουμε τον αριθμό των κλάσεων K είτε αυθαίρετα είτε βάση κάποιου κανόνα. Για παράδειγμα $K = \lceil\sqrt{N}\rceil$ ή $K = 1 + \lceil 3.322 \log N\rceil$ (κανόνας του Sturges). Το μέγεθος κάθε κλάσης δίνεται από τον τύπο

$$d = \frac{R}{K}\,, \tag{1}$$

όπου R το *εύρος* των δεδομένων,

$$R = M - m\,, \tag{2}$$

και $M = \max\{x_i : i = 1, \ldots, N\}$ και $m = \min\{x_i : i = 1, \ldots, N\}$. Ορίζουμε το διάστημα της κλάσης C_i ως

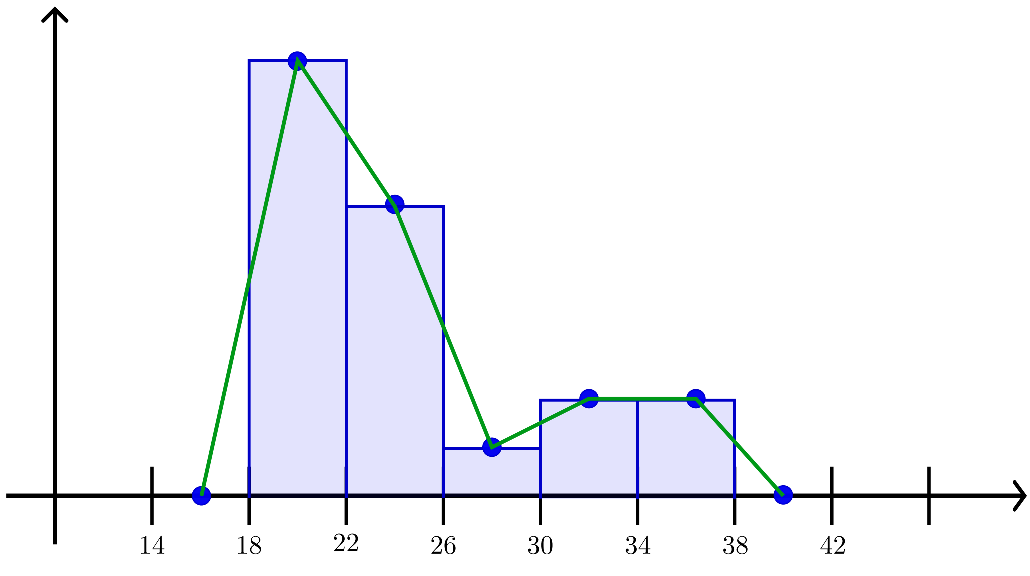
$$C_k : \quad \big[m + (k-1)d,\, m + kd\big), \quad k = 1, \ldots, K\,. \tag{3}$$

class	frequency	relative freq.	cumulative freq.	relative cumulative freq.
C_1	$f_1 = 9$	$rf_1 = \frac{f_1}{N} = \frac{9}{20}$	$F_1 = f_1 = 9$	$rF_1 = \frac{f_1}{N} = \frac{9}{20}$
C_2	$f_2 = 6$	$rf_2 = \frac{f_2}{N} = \frac{6}{20}$	$F_2 = F_1 + f_2 = 15$	$rF_2 = \frac{f_2}{N} = \frac{15}{20}$
C_3	$f_3 = 1$	$rf_3 = \frac{f_3}{N} = \frac{1}{20}$	$F_3 = F_2 + f_3 = 16$	$rF_3 = \frac{f_3}{N} = \frac{16}{20}$
C_4	$f_4 = 2$	$rf_4 = \frac{f_4}{N} = \frac{2}{20}$	$F_4 = F_3 + f_4 = 18$	$rF_4 = \frac{f_4}{N} = \frac{18}{20}$
C_5	$f_5 = 2$	$rf_5 = \frac{f_5}{N} = \frac{2}{20}$	$F_5 = F_4 + f_5 = 20 = N$	$rF_5 = \frac{F_5}{N} = 1$

Πίνακας 3: Πίνακας συχνοτήτων για τα αριθμητικά δεδομένα του Πίνακα 2.

Ορισμός 1.4 (συχνότητες)**.** *Η συχνότητα (frequency) f ορίζεται ως ο αριθμός των δεδομένων που ανήκουν σε κάθε κλάση. Η σχετική συχνότητα (relative frequency) $\hat{f}$ ορίζεται ως η συχνότητα διαιρεμένη με το μέγεθος του συνόλου δεδομένων N. Η αθροιστική συχνότητα (cumulative frequency) της κλάσης C_i ορίζεται ως $F_i = \sum_{j=1}^{i} f_j$. Η σχετική αθροιστική συχνότητα (relative cumulative frequency) ορίζεται ως $\hat{F}_i = F_i/N$.*

Στον Πίνακα 3 παρουσιάζεται ο πίνακας των συχνοτήτων.



Σχήμα 1: Ιστόγραμμα σχετικών συχνοτήτων χρησιμοποιώντας τα δεδομένα του Πίνακα 3.

1.4 Οργάνωση ποιοτικών δεδομένων

Τα ποιοτικά δεδομένα μπορούν να οργανωθούν σε πίνακα συχνοτήτων θεωρώντας κάθε δυνατή τιμή μιας μεταβλητής ώς μια κλάση. Με τον ίδιο τρόπο όπως τα ποσοτικά δεδομένα μπορούμε να σχεδιάσουμε το ιστόγραμμα.

Ένα ακόμη διάγραμμα για την παρουσίαση των δεδομένων είναι το διάγραμμα πίτας. Χωρίζουμε έναν κύκλο σε τόσα κομμάτια όσα ο αριθμός των κλάσεων. Η γωνία την κάθε κλάσης δίνεται ως $\omega_i = 360\hat{f}_i$.

2 Εβδομάδα 02/13

2.1 Περιγραφικά μέτρα: Τυχαίες μεταβλητές

Οι βασικές κατηγορίες περιγραφικών μέτρων είναι οι εξής:

1. *Μέτρα θέσης:* τιμή γύρω από την οποία συγκεντρώνονται οι τιμές της μεταβλητής.
2. *Μέτρα μεταβλητότητας:* ποσοτικοποιούν πόσο μακριά απλώνονται οι τιμές της μεταβλητής από κάποιο μέτρο κεντρικής τάσης.
3. *Μέτρα ασυμμετρίας:* ποσοτικοποιούν την συμμετρία γύρω από κάποιο μέτρο κεντρικής τάσης.
4. *Μέτρα κύρτωσης:* ποσοτικοποιούν την οξύτητα της κορυφής της κατανομής των τιμών της μεταβλητής.

2.2 Μέτρα θέσης

2.2.1 Μέση τιμή

Ορισμός 2.1 (μέση τιμή). *Η μέση τιμή (mean) $\{x_i\}_{i=1}^N$ ορίζεται ως,*

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \,. \tag{4}$$

Η μέση τιμή του πληθυσμού (population mean) είναι η μέση τιμή όλων των δυνατών τιμών των στοιχείων ενός πληθυσμού και συμβολίζεται με μ. Η μέση τιμή του δείγματος (sample mean) είναι η μέση τιμή όλων των δυνατών τιμών των στοιχείων ενός υποσυνόλου του πληθυσμού και συμβολίζεται με $\bar{X}$.

Εάν ο πληθυσμός περιέχει άπειρα στοιχεία, τότε ισχύει ότι $\bar{X} \to \mu$ καθώς $N \to \infty$.

Άσκηση 2.1. *Δείξτε ότι αν $y_i = ax_i + b$ τότε $\bar{Y} = a\bar{X} + b$.*

Ορισμός 2.2 (σταθμισμένη μέση τιμή). *Αν κάθε τιμή x_i έχει ένα βάρος w_i, η σταθμισμένη μέση τιμή ορίζεται ως,*

$$\bar{X} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i} \,. \tag{5}$$

Για ομαδοποιημένες παρατηρήσεις σε κλάσεις C_i, $i = 1, \ldots, M$, με συχνότητα κάθε κλάσης f_i και το μέσο κάθε κλάσης m_i, η μέση τιμή μπορεί να προσεγγιστεί από την τιμή,

$$\bar{X} \approx \frac{1}{M}\sum_{i=1}^{M} f_i m_i \,, \tag{6}$$

ή ισοδύναμα,

$$\bar{X} \approx \sum_{i=1}^{M} \hat{f}_i m_i \,. \tag{7}$$

Παρατηρήστε ότι η προσέγγιση είναι μια σταθμισμένη μέση τιμή των μέσων των κλάσεων, με βάρη $w_i = f_i$.

2.2.2 Διάμεσος

Ορισμός 2.3 (διάμεσος). *Διάμεσος (median) είναι η τιμή που χωρίζει τις τιμές της μεταβλητής με τέτοιο τρόπο ώστε το 50% αυτών να είναι μικρότερες ή ίσες από την διάμεσο,*

$$m = \begin{cases} x_{\frac{N+1}{2}} & N \text{ περιττός} \\ \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}) & N \text{άρτιος}\,. \end{cases} \tag{8}$$

Για ομαδοποιημένες παρατηρήσεις, με διάστημα κάθε κλάσης $[a_i, a_{i+1})$, $a_{i+1} - a_i = d$, $i = 1, \ldots, M$, συχνότητα f_i, και αθροιστική συχνότητα F_i, υποθέτουμε ότι υπάρχει μοναδικός δείκτης j τέτοιος ώστε

$$F_{j-1} < \frac{N}{2} \leq F_j \,. \tag{9}$$

Ορίζουμε ως προσέγγιση της διάμεσου την τιμή,

$$m = a_j + d\frac{\frac{N}{2} - F_{j-1}}{f_j}\,. \tag{10}$$

Παρατηρήστε ότι αν $F_j = \frac{N}{2}$ τότε $m = a_j + d = a_{j+1}$, και εάν $F_{j-1} = \frac{N}{2}$ τότε $m = a_j$.

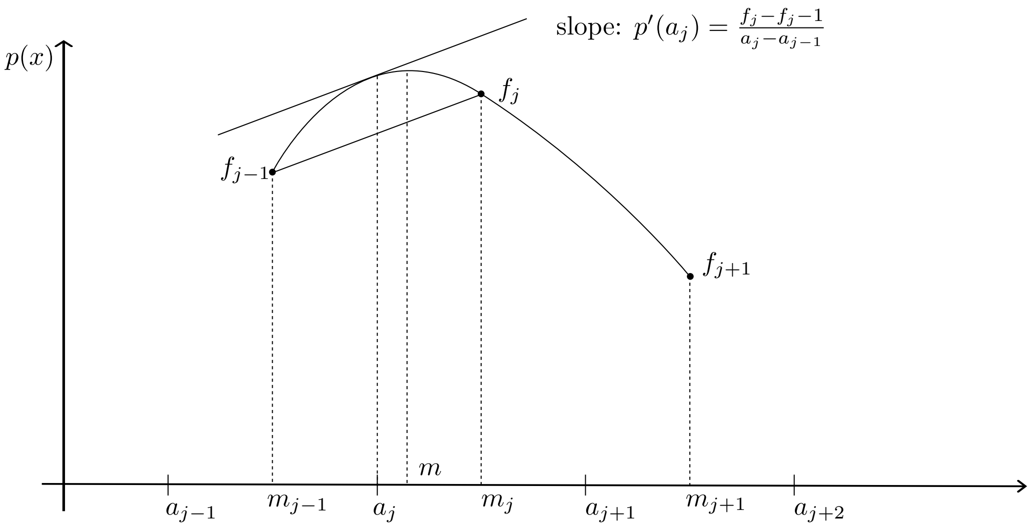
2.2.3 Επικρατέστερη τιμή

Ορισμός 2.4 (επικρατέστερη τιμή). *Επικρατέστερη τιμή (mode) M είναι η τιμή της μεταβλητής που εμφανίζεται με την μεγαλύτερη συχνότητα. Ορίζεται και για ποιοτικές μεταβλητές. Εάν δύο ή περισσότερες τιμές έχουν την ίδια μέγιστη συχνότητα τότε δεν ορίζεται επικρατέστερη τιμή.*

Για ομαδοποιημένες παρατηρήσεις, εάν υπάρχει δείκτης j τέτοιος ώστε $f_j > f_i$, $i \neq j$, τότε $M \in [a_j, a_{j+1})$ και μια προσέγγιση της επικρατέστερης τιμής δίνεται από τη σχέση

$$M \approx a_j + d\frac{f_j - f_{j-1}}{(f_j - f_{j-1}) - (f_{j+1} - f_j)}\,. \tag{11}$$

Η προσέγγιση στηρίζεται στην ακόλουθη μεθοδολογία. Θεωρούμε τις κλάσεις C_{j-1}, C_j, C_{j+1}, με μέσο m_{j-1}, m_j, m_{j+1} και συχνότητες f_{j-1}, f_j, f_{j+1}. Από τα τρία αυτά σημεία παρεμβάλλεται μοναδικό πολυώνυμο δεύτερου βαθμού, p (δείτε το Σχήμα 2). Θα προσεγγίσουμε την επικρατέστερη τιμή από το σημείο στο οποίο παίρνει μέγιστο το πολυώνυμο.



Σχήμα 2: Στα σημεία $(m_{j-1}, f_{j-1}), (m_j, f_j), (m_{j+1}, f_{j+1})$ παρεμβάλουμε ένα πολυώνυμο δευτέρου βαθμού, p. Υποθέτουμε ότι η p έχει μέγιστο στο $m \in [a_{j-1}, a_j]$.

Αρχικά παρατηρούμε ότι

$$p'(a_j) = \frac{f_j - f_{j-1}}{m_j - m_{j-1}} = \frac{f_j - f_{j-1}}{d}\,, \tag{12}$$

και

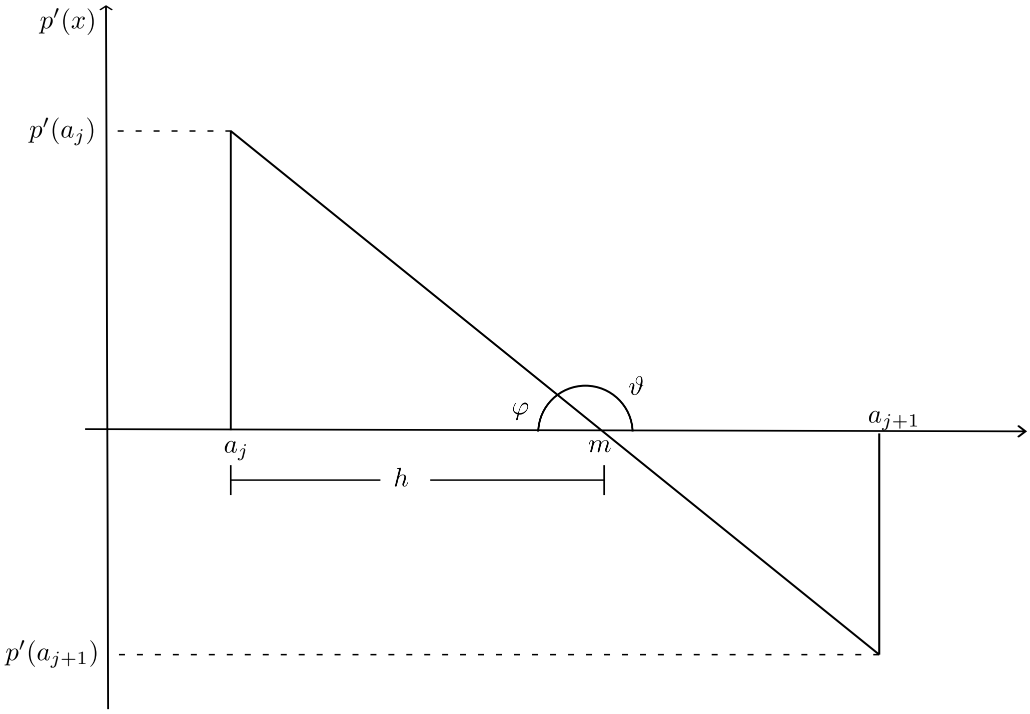
$$p'(a_{j+1}) = \frac{f_{j+1} - f_j}{m_{j+1} - m_j} = \frac{f_{j+1} - f_j}{d}\,. \tag{13}$$

Οι σχέσεις αυτές είναι αληθής διότι αν p είναι πολυώνυμο δεύτερο βαθμού, τότε για κάθε διάστημα (c, d) ισχύει,

$$p'(\frac{c+d}{2}) = \frac{p(d) - p(c)}{d - c}\,. \tag{14}$$

Η πρώτη παράγωγος του p είναι γραμμική συνάρτηση και την δείχνουμε στο Σχήμα 3. Συμβολίζουμε το σημείο m που μηδενίζεται η πρώτη παράγωγος με m και ψάχνουμε να βρούμε την τιμή του $h = m - a_j$. Από το Σχήμα 3 παρατηρούμε ότι

$$\tan\varphi = \frac{p'(a_j)}{h} \Rightarrow h = \frac{p'(a_j)}{\tan\varphi}\,. \tag{15}$$



Σχήμα 3: Η πρώτη παράγωγος ενός πολυώνυμου δευτέρου βαθμού είναι γραμμική συνάρτηση. Η παράγωγος μηδενίζεται στο m και ο άγνωστος είναι η απόσταση h από το άκρο a_j της κλάσης C_j.

Το $p'(a_j)$ μας είναι γνωστό από την σχέση (12). Παρατηρήστε ότι για το $\tan\varphi$ ισχύει ότι,

$$\tan\varphi = \tan(\pi - \vartheta) = -\tan\vartheta = -\frac{p'(a_{j+1}) - p'(a_j)}{d}\,. \tag{16}$$

Τέλος, έχουμε ότι,

$$m = a_j + h = a_j + d\frac{f_j - f_{j-1}}{(f_j - f_{j-1}) - (f_{j+1} - f_j)}\,. \tag{17}$$

2.3 Μέτρα μεταβλητότητας

2.3.1 p-ποσοστημόρια

Ορισμός 2.5 (ποσοστημόρια). *Για $p \in (0,1)$ ορίζουμε το p-ποσοστημόριο του δείγματος ως την τιμή P_p που χωρίζει τις τιμές της μεταβλητής με τέτοιο τρόπο ώστε το $100p\%$ αυτών να είναι μικρότερες ή ίσες από την P_p. Για $p = 0.5$ παίρνουμε την διάμεσο.*

Για ένα διατεταγμένο σύνολο $\{x_i\}_{i=1}^N$ με $x_i < x_{i+1}$, $i = 1, \ldots, N-1$,

$$P_p = \begin{cases} x_{p(N-1)+1} & p(N-1) \in \mathbb{Z} \\ x_{\lceil p(N-1)\rceil} + u\big(x_{\lceil p(N-1)+1\rceil} - x_{\lceil p(N-1)\rceil}\big) & p(N-1) \notin \mathbb{Z} \end{cases}, \tag{18}$$

όπου $u = p(N-1) - \lfloor p(N-1)\rfloor$, το δεκαδικό μέρος του $p(N-1)$. Η συνάρτηση $\lfloor x\rfloor$ είναι ο μεγαλύτερος ακέραιος που είναι μικρότερος από το x, και $\lceil x\rceil$ είναι ο μικρότερος ακέραιος που είναι μεγαλύτερος από το x.

Παράδειγμα 2.1. *Να βρεθεί το 35% ποσοστημόριο των παρατηρήσεων $\{3, 4, 7, 10, 12, 17\}$. Γι αυτό το σύνολο έχουμε $N = 6$, και $P(N-1) = 0.35(6-1) = 1.75 \notin \mathbb{Z}$. Το ποσοστημόριο δίνεται από το δεύτερο σκέλος της σχέσης (18). Υπολογίζουμε $\lfloor p(N-1) = 1\rfloor$, $u = 1.75 - 1 = 0.75$, και*

$$P_{0.35} = x_2 + u(x_3 - x_2) = 4 + 0.75(7-4) = 6.25\,.$$

2.3.2 Τεταρτημόρια

Ορισμός 2.6 (τεταρτημόρια). *Ορίζουμε τρία τεταρτημόρια (quartiles), $Q_1 = P_{0.25}$, $Q_2 = P_{0.5}$, $Q_3 = P_{0.75}$.*

Για ομαδοποιημένες παρατηρήσεις δουλεύουμε όπως και στην Παράγραφο 2.2.2. Για $q = 1, 2, 3$ υποθέτουμε ότι υπάρχει μοναδικός δείκτης j τέτοιος ώστε

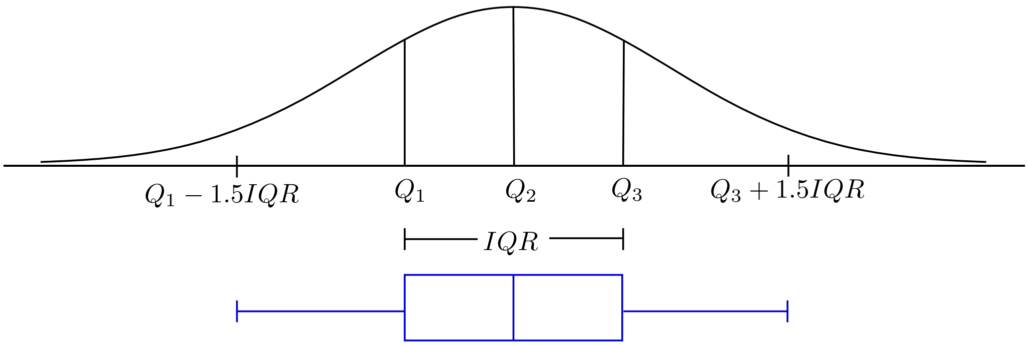
$$F_{j-1} < q\frac{N}{4} \leq F_j \,, \tag{19}$$

και ορίζουμε την προσέγγιση του Q_q ως,

$$Q_q = a_j + d\frac{q\frac{N}{4} - F_{j-1}}{f_j} \,. \tag{20}$$

Ορισμός 2.7 (ενδοτεταρτημοριακό εύρος)**.** *Την απόσταση $IQR = Q_3 - Q_1$ την ονομάζουμε* ενδοτεταρτημοριακό εύρος *(interquartile range).*

2.3.3 Box plot



Σχήμα 4: Box plot.

2.3.4 Διασπορά

Ορισμός 2.8 (διασπορά)**.** *Η διασπορά του πληθυσμού $\{x_i\}_{i=1}^N$ ορίζεται ως*

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \,. \tag{21}$$

Η διασπορά του στατιστικού δείγματος $\{x_i\}_{i=1}^N$ ορίζεται ως

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2 \,. \tag{22}$$

Εάν ο πληθυσμός περιέχει άπειρα στοιχεία, τότε ισχύει ότι $s^2 \to \sigma^2$ καθώς $N \to \infty$.

Για ομαδοποιημένες παρατηρήσεις σε κλάσεις C_i, $i = 1, \ldots, M$, με σχετική συχνότητα κάθε κλάσης $\hat{f}_i$ και το μέσο κάθε κλάσης m_i, η διακύμανση μπορεί να προσεγγιστεί από την τιμή,

$$s^2 = \sum_{i=1}^{N} \hat{f}_i (m_i - \bar{X})^2 \,, \tag{23}$$

όπου $\bar{X} = \sum_{i=1}^N \hat{f}_i m_i$, βλέπε σχέση (7).

Η τυπική απόκλιση του πληθυσμού ορίζεται ως $\sigma = \sqrt{\sigma^2}$ και του δείγματος ως $s = \sqrt{s^2}$. Παρατηρήστε ότι η τυπική απόκλιση έχει τις ίδιες μονάδες με την μέση τιμή.

Θεώρημα 2.1 (Chebyshev)**.** *Για $k > 0$ τουλάχιστον $(1 - \frac{1}{k^2})100\%$ των στοιχείων ενός πληθυσμού ανήκουν στο διάστημα $[\mu - k\sigma, \mu + k\sigma]$.*

Το θεώρημα μπορεί να χρησιμοποιηθεί σε ένα στατιστικό δείγμα αντικαθιστώντας την μέση τιμή του πληθυσμού με $\bar{X}$ και την τυπική απόκλιση του πληθυσμού με s.

Παράδειγμα 2.2. *Η μέση τιμή της συστολικής αρτηριακής πίεσης 400 ατόμων βρέθηκε να είναι 187 με τυπική απόκλιση 22. Τουλάχιστον πόσα άτομα έχουν πίεση μεταξύ 143 και 231; Χρησιμοποιούμε το θεώρημα του Chebyshev για $k = 2$ και βρίσκουμε ότι το 75% του δείγματος*

Ορισμός 2.9 (συντελεστής μεταβλητότητας). *Ο συντελεστής μεταβλητότητας (coefficient of variation) ορίζεται ως*

$$CV = \frac{\sigma}{\mu}\,. \tag{24}$$

Χρησιμοποιείται ως μέτρο σύγκρισης της μεταβλητότητας μεταξύ πληθυσμών που παίρνουν τιμές σε διαφορετικές κλίμακες.

3 Εβδομάδα 03/13

3.1 Μέτρα ασυμμετρίας

Ορισμός 3.1 (Pearson). *Ο συντελεστής ασυμμετρίας του Pearson για έναν πληθυσμό με μέση τιμή μ, επικρατέστερη τιμή M, και τυπική απόκλιση σ ορίζεται ως*

$$S_{kp} = \frac{\mu - M}{\sigma}\,. \tag{25}$$

Εάν είναι διαθέσιμο μόνο ένα δείγμα από τον πληθυσμό, αντικαθιστούμε την μέση τιμή, επικρατέστερη τιμή, και τυπική απόκλιση, με τις εκτιμήσεις τους.

Ορισμός 3.2 (Bowley). *Ο συντελεστής ασυμμετρίας του Bowley για έναν πληθυσμό με τεταρτημόρια $Q_1, Q2$, και $Q3$, ορίζεται ως*

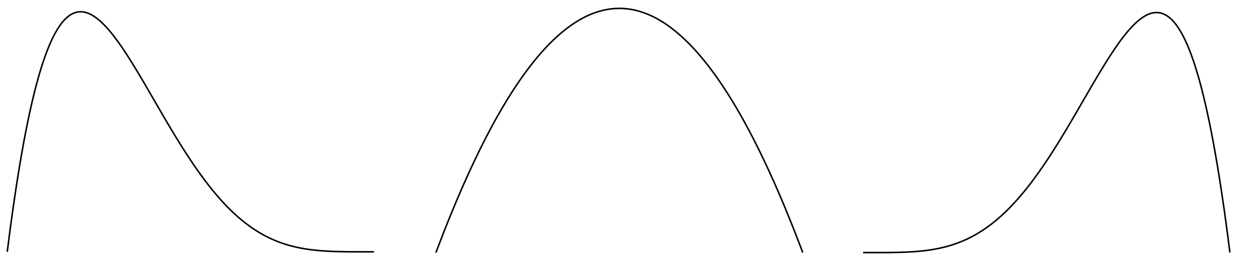
$$S_b = \frac{\frac{Q_1+Q_3}{2} - Q_2}{\frac{Q_1+Q_3}{2}} = 1 - \frac{Q_2}{\frac{Q_1+Q_3}{2}}\,. \tag{26}$$

Ορισμός 3.3 (Kelly). *Ο συντελεστής ασυμμετρίας του Kelly για έναν πληθυσμό με ποσοστημόρια $P_{0.1}, P_{0.5}$, και $P_{0.9}$, ορίζεται ως*

$$S_b = 1 - \frac{P_{0.5}}{\frac{P_{0.1}+P_{0.9}}{2}}\,. \tag{27}$$

Ποιο γενικά, για $p \in (0, 0.5)$,

$$S_b = 1 - \frac{P_{0.5}}{\frac{P_p+P_{1-p}}{2}}\,. \tag{28}$$



Σχήμα 5: Κατανομές με θετική ασυμμετρία, συμμετρία, και αρνητική ασυμμετρία.

3.2 Καμπύλη Lorenz

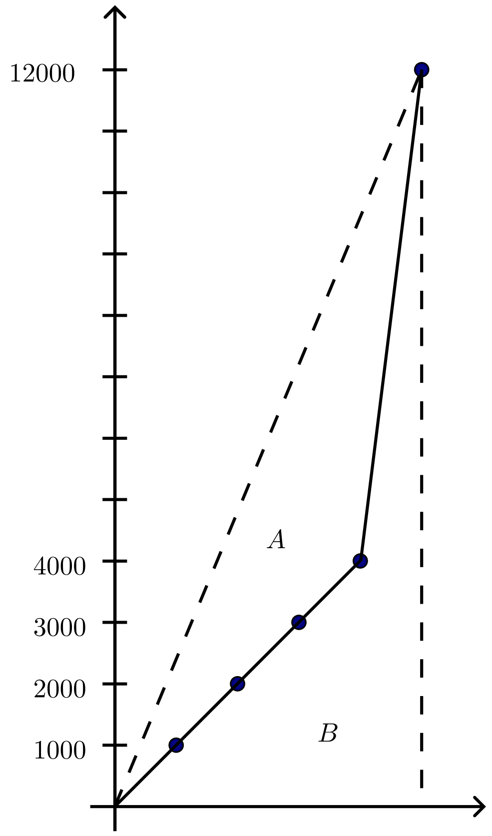
3.3 Πιθανότητες

Ορισμός 3.4 (Δειγματικός χώρος). Δειγματικός χώρος Ω *είναι το σύνολο όλων των δυνατών αποτελεσμάτων ενός πειράματος. Τα στοιχεία $\omega \in \Omega$ ονομάζονται* δείγματα *και τα υποσύνολα του Ω ενδεχόμενα.*

Παράδειγμα 3.1. *Ρίχνουμε ένα κέρμα δύο φορές, $\Omega = \{KK, K\Gamma, \Gamma K, \Gamma\Gamma\}$. Το ενδεχόμενο το πρώτο γράμμα να είναι K είναι το υποσύνολο $\{KK, K\Gamma\}$*

Παράδειγμα 3.2. *Ρίχνουμε ένα κέρμα δέκα φορές, $\Omega = \{\omega = (\omega_1, \omega_2, \ldots, \omega_{10}) : \omega_i \in \{K, \Gamma\},\ i = 1, 2, \ldots, 10\}$. Το ενδεχόμενο να εμφανιστεί Γ για πρώτη φορά στην τρίτη ρίψη είναι το σύνολο $A = \{\omega = (\omega_1, \omega_2, \ldots, \omega_{10}) : \omega_1 = \omega_2 = K, \omega_3 = \Gamma, \omega_i \in \{K, \Gamma\},\ i = 4, 5, \ldots, 10\}$.*

Αν A, B δύο ενδεχόμενα, τότε $A \cup B = \{\omega \in \Omega : \omega \in A, \omega \notin B \text{ ή } \omega \in B, \omega \notin A \text{ ή } \omega \in A, \omega \in B\}$ είναι το ενδεχόμενο να συμβεί είτε το A, είτε το B, είτε και τα δύο μαζί, και $A \cap B = \{\omega \in \Omega : \omega \in A \text{ και } \omega \in B\}$ είναι το ενδεχόμενο να συμβούν και τα δύο μαζί.



Σχήμα 6: Διάγραμμα Lorenz για τα δεδομένα $\{1000, 1000, 1000, 1000, 8000\}$. Αν υποθέσουμε ότι τα δεδομένα αντιστοιχούν σε μισθούς 5 υπαλλήλων, ο λόγος των εμβαδών $\frac{A}{A+B}$ είναι ένα μέτρο της ανισοκατανομής των τιμών. Το μέτρο ονομάζεται συντελεστής Gini.

Ορισμός 3.5 (κατανομή πιθανότητας). *Ορίζουμε την* πιθανότητα *ως μια συνάρτηση* $\mathbb{P}$ *που δίνει πραγματικές τιμές στα ενδεχόμενα, αν ικανοποιεί:*

- $\mathbb{P}(A) \geq 0$ *για κάθε* $A \subset \Omega$.
- $\mathbb{P}(\Omega) = 1$.
- $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} A_i$ *όταν τα* A_i *είναι ξένα ανά δύο.*

Η συνάρτηση ονομάζεται κατανομή πιθανότητας *ή* μέτρο πιθανότητας.

Για ένα ενδεχόμενο A ορίζουμε A^c ως το ενδεχόμενο να μη συμβεί το A, δηλαδή $A^c = \Omega \setminus A$

Άσκηση 3.1. *Δείξτε ότι* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

3.4 Πιθανότητα σε πεπερασμένους δειγματικούς χώρους

Αν $|\Omega| < \infty$ και κάθε αποτέλεσμα του πειράματος είναι το ίδιο πιθανό να συμβεί, τότε ορίζουμε

$$P(A) = \frac{|A|}{|\Omega|}. \tag{29}$$

Αυτή την κατανομή πιθανότητας την ονομάζουμε *ομοιόμορφη κατανομή*.

Παράδειγμα 3.3. *Ρίχνουμε ένα ζάρι 2 φορές. Ποια είναι η πιθανότητα το άθροισμα να είναι ίσο με 11;* $\Omega = \{(i,j) : i,j \in \{1,2,\ldots,6\}\}$, $A = \{(5,6),(6,5)\}$, *και* $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{36} = \frac{1}{18}$.

Για να βρούμε το μέγεθος ενός ενδεχομένου πρέπει να μετρήσουμε όλα τα στοιχεία του συνόλου. Υπάρχουν 3 βασικοί τρόποι για να μετρήσουμε.

- Όλοι οι τρόποι που μπορούν να μπουν n αντικείμενα σε σειρά είναι $n! = n(n-1)\ldots 2 \cdot 1$
- Από n αντικείμενα διαλέγουμε k και τα βάζω σε σειρά, $n(n-1)\ldots(n-k-1) = \frac{n!}{k!}$

- Από n αντικείμενα διαλέγουμε k χωρίς να μας ενδιαφέρει η σειρά $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

3.5 Ανεξάρτητα ενδεχόμενα

Λέμε ότι δύο ενδεχόμενα είναι ανεξάρτητα όταν $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Παράδειγμα 3.4. *Ρίχνουμε ένα κέρμα 10 φορές. Έστω A το ενδεχόμενο να έρθει τουλάχιστον μία φορά K. Ποια είναι η πιθανότητα του A; Η άρνηση αυτού του ενδεχομένου είναι να μην έρθει καμία φορά K, διαφορετικά, να έρθει Γ σε κάθε ρίψη. Έστω T_i το ενδεχόμενο να έρθει Γ στην i ρίψη. Παρατηρήστε ότι το ενδεχόμενο να έρθει Γ σε κάθε ρίψη είναι το σύνολο $T = \bigcap_{i=1}^{10} T_i$ και ότι τα ενδεχόμενα T_i είναι ανεξάρτητα. Τότε $\mathbb{P}(T) = \prod_{i=1}^{10} \mathbb{P}(T_i) = 2^{-10}$. Τέλος $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - 2^{-10}$.*

3.6 Τυχαίες μεταβλητές

Ορισμός 3.6 (τυχαία μεταβλητή)**.** *Μία* τυχαία μεταβλητή *είναι μια απεικόνιση $X : \Omega \to \mathbb{R}$. Δοσμένου $A \subset \mathbb{R}$, ορίζουμε $X^{-1}(A) = \{\omega \in \Omega : X(\omega) = x\}$ και*

$$\mathbb{P}_X(X \in A) = \mathbb{P}\big(X^{-1}(A)\big) = \mathbb{P}\big(\{\omega \in \Omega : X(\omega) = x\}\big)\,. \tag{30}$$

Παράδειγμα 3.5. *Ρίχνουμε ένα νόμισμα 3 φορές. Ορίζουμε $X(\omega)$ να είναι ο αριθμός των Γ στο ω. Για παράδειγμα αν $\omega = K\Gamma\Gamma$, τότε $X(\omega) = 2$ και η $\mathbb{P}_X$ δίνεται από τον Πίνακα 4.*

ω	$\mathbb{P}(\omega)$	$X(\omega)$
KKK	1/8	0
$KK\Gamma$	1/8	1
$K\Gamma K$	1/8	1
$K\Gamma\Gamma$	1/8	2
ΓKK	1/8	1
$\Gamma K\Gamma$	1/8	2
$\Gamma\Gamma K$	1/8	2
$\Gamma\Gamma\Gamma$	1/8	3

x	$\mathbb{P}(x)$
0	1/8
1	3/8
2	3/8
3	1/8

Πίνακας 4: Ρίψη νομίσματος 3 φορές και ο αριθμός των Γ στο αποτέλεσμα.

3.6.1 Διακριτές τυχαίες μεταβλητές

Ορισμός 3.7 (διακριτή τυχαία μεταβλητή)**.** *Μια τυχαία μεταβλητή είναι* διακριτή *όταν παίρνει αριθμήσιμες το πλήθος τιμές.*

Ορισμός 3.8 (συνάρτηση κατανομής)**.** *Ορίζουμε την* συνάρτηση κατανομής *ή* αθροιστική συνάρτηση κατανομής *(cumulative distribution function, CDF), ως την συνάρτηση $F_X : \mathbb{R} \to [0,1]$, που δίνεται από την σχέση,*

$$F_X(x) = \mathbb{P}_X(X \le x)\,. \tag{31}$$

Παράδειγμα 3.6. *Ρίχνουμε ένα νόμισμα 3 φορές. Από τον Πίνακα 4,*

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}, & 0 \le x < 1 \\ \frac{4}{8}, & 1 \le x < 2 \\ \frac{7}{8}, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases} \tag{32}$$

Ορισμός 3.9 (συνάρτηση μάζας πιθανότητας)**.** *Ορίζουμε την* συνάρτηση πιθανότητας *ή* συνάρτηση μάζας πιθανότητας *(probability mass function, PMF) για μια διακριτή τυχαία μεταβλητή ως,*

$$f_X(x) = \mathbb{P}(X = x)\,. \tag{33}$$

Παρατηρήστε ότι $f_X \geq 0$, $\sum_i f_X(x_i) = 1$, και ότι

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)\,. \tag{34}$$

Παράδειγμα 3.7. *Ρίχνουμε ένα νόμισμα 3 φορές. Από τον Πίνακα 4,*

$$f_X(x) = \begin{cases} \frac{1}{8}, & x = 0 \\ \frac{3}{8}, & x = 1 \\ \frac{3}{8}, & x = 2 \\ \frac{1}{8}, & x = 3 \end{cases} \tag{35}$$

3.6.2 Συνεχείς τυχαίες μεταβλητές

Ορισμός 3.10 (συνάρτηση πυκνότητας πιθανότητας)**.** *Μία τυχαία μεταβλητή είναι συνεχής αν υπάρχει συνάρτηση f_X τέτοια ώστε $f_X(x) \geq 0\,, \forall x \in \mathbb{R}$, $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$, και για κάθε $a < b$,*

$$\mathbb{P}_X(X \in (a,b)) = \int_a^b f_X(x)\,\mathrm{d}x\,. \tag{36}$$

Η συνάρτηση f_X ονομάζεται συνάρτηση πυκνότητας πιθανότητας *(probability density function, PDF).*

Ορισμός 3.11 (συνάρτηση κατανομής)**.** *Η συνάρτηση κατανομής δίνεται ως,*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t\,, \tag{37}$$

και για όλα τα σημεία στα οποία η F_X είναι παραγωγίσιμη, ισχύει ότι $f_X(x) = F_X'(x)$.

Ερώτηση 3.1. *Είναι η*

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{1+x}, & x \geq \end{cases}, \tag{38}$$

συνάρτηση πυκνότητας πιθανότητας;

Παρατήρηση 3.1. *Για τις συνεχείς τυχαίες μεταβλητές ισχύει ότι $\mathbb{P}_X(X = x) = 0$. Είναι λάθος να σκεφτόμαστε την f_X σαν την πιθανότητα $\mathbb{P}_X$. Αυτό ισχύει μόνο για διακριτές τυχαίες μεταβλητές. Για να βρούμε πιθανότητες πρέπει να ολοκληρώσουμε. Τέλος, μια συνάρτηση πυκνότητας πιθανότητας μπορεί να είναι μεγαλύτερη της μονάδας, σε αντίθεση με την συνάρτηση μάζας πιθανότητας. Η συνάρτηση πυκνότητας πιθανότητας μπορεί να είναι μη φραγμένη, αρκεί το ολοκλήρωμα της να είναι 1.*

4 Εβδομάδα 04/13

4.1 Συνεχείς τυχαίες μεταβλητές

Ορισμός 4.1 (συνάρτηση ποσοστημόριου)**.** *Η αντίστροφη συνάρτηση πυκνότητας πιθανότητας ή* συνάρτηση ποσοστημόριου *(quartile function) ορίζεται ως*

$$F^{-1}(q) = \inf\{x : F(x) > q\}\,. \tag{39}$$

Εάν η F είναι γνησίως αύξουσα και συνεχής, τότε η τιμή της $F^{-1}(q)$ είναι εκείνο το x για το οποίο ισχύει $F(x) = q$.

Γράφουμε $X \sim F$ όταν η X ακολουθεί την συνάρτηση κατανομής F.

Λέμε ότι η X ακολουθεί την *ομοιόμορφη κατανομή* όταν η X έχει συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}, \tag{40}$$

και γράφουμε $X \sim \mathcal{U}(a,b)$.

Άσκηση 4.1. *Βρείτε την συνάρτηση κατανομής της ομοιόμορφης κατανομής.*

Λέμε ότι η X ακολουθεί *κανονική κατανομή* με παραμέτρους μ και σ όταν

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R}\,, \tag{41}$$

και γράφουμε $X \sim \mathcal{N}(\mu, \sigma^2)$. Για $\mu = 0$ και $\sigma = 1$ η κατανομή λέγεται *τυπική κανονική κατανομή*. Σ' αυτή την περίπτωση συμβολίζουμε την τυχαία μεταβλητή με Z, την συνάρτηση πυκνότητας με ϕ, και την συνάρτηση κατανομής με Φ.

Άσκηση 4.2. *Δείξτε ότι αν* $X \sim \mathcal{N}(\mu, \sigma^2)$ *τότε η* $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Άσκηση 4.3. *Δείξτε ότι αν* $Z \sim \mathcal{N}(0, 1)$ *τότε η* $X = \mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)$.

Για να αποδείξουμε τις παραπάνω σχέσεις χρειαζόμαστε το *θεώρημα αλλαγής μεταβλητής*.

Θεώρημα 4.1 (αλλαγή μεταβλητής)**.** *Έστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας f_X συνεχή και g γνησίως μονότονη συνάρτηση. Η συνάρτηση πυκνότητας πιθανότητας, f_Y, για την μεταβλητή $Y = g(X)$, δίνεται από την σχέση,*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} \right| . \tag{42}$$

Απόδειξη. Η f_Y θα πρέπει να δίνει την ίδια πιθανότητα με την f_X στα αντίστοιχα διαστήματα, δηλαδή,

$$\mathbb{P}_Y\left(Y \in I\right) = \mathbb{P}_X\left(X \in g^{-1}(I)\right)\,, \tag{43}$$

για κάθε $I = (a, b)$ διάστημα στο $\mathbb{R}$. Η μονοτονία της g χρειάζεται για να υπάρχει η g^{-1}. Αρχικά υποθέτουμε ότι η g είναι αύξουσα. Από την εξίσωση (43) έχουμε,

$$\int_a^b f_Y(y)\,\mathrm{d}y = \int_{g^{-1}(a)}^{g^{-1}(b)} f_X(x)\,\mathrm{d}x\,, \tag{44}$$

και κάνουμε την αλλαγή μεταβλητής $x = g^{-1}(y)$ όπου $\mathrm{d}x = \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\,\mathrm{d}y$

$$\int_a^b f_Y(y)\,\mathrm{d}y = \int_a^b f_X(g^{-1}(y)) \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\,\mathrm{d}y\,. \tag{45}$$

Η πάνω σχέση ισχύει για κάθε $(a, b) \subset \mathbb{R}$, επομένως,

$$f_Y(y) = f_X(g^{-1}(y)) \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\,. \tag{46}$$

Εάν η g είναι γνησίως φθίνουσα, τότε το διάστημα (a, b) απεικονίζεται μέσω της g^{-1} στο διάστημα $(g^{-1}(b), g^{-1}(a))$. Επομένως η εξίσωση (44) γράφεται ως

$$\int_a^b f_Y(y)\,\mathrm{d}y = \int_{g^{-1}(b)}^{g^{-1}(a)} f_X(x)\,\mathrm{d}x\,, \tag{47}$$

Κάνοντας της αλλαγή μεταβλητής και λαμβάνοντας υπόψιν ότι η παράγωγος της g είναι αρνητική έχουμε

$$\int_a^b f_Y(y)\,\mathrm{d}y = \int_b^a f_X(g^{-1}(y)) \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\,\mathrm{d}y = -\int_a^b f_X(g^{-1}(y)) \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\,\mathrm{d}y = \int_a^b f_X(g^{-1}(y)) \left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right| \mathrm{d}y\,. \tag{48}$$

Από τις σχέσεις (46) και (48) προκύπτει το ζητούμενο. □

Άσκηση 4.4. *Εάν f συνεχής και $\int_a^b f(x)\,\mathrm{d}x = 0$ για κάθε $(a, b) \subset \mathbb{R}$ δείξτε ότι $f \equiv 0$.*

Παράδειγμα 4.1. *Για την τυχαία μεταβλητή $X \sim \mathcal{N}(3, 5)$ βρείτε την τιμή $\mathbb{P}(X > 1)$ και την τιμή q για την οποία ισχύει $\mathbb{P}(X < q) = 0.2$.*

$$\mathbb{P}_X\left(X > 1\right) = 1 - \mathbb{P}_X\left(X < 1\right) = 1 - \mathbb{P}_Z\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-0.8944) = 0.81. \tag{49}$$

$$0.2 = \mathbb{P}_X\left(X < q\right) = \mathbb{P}_Z\left(Z < \frac{q-\mu}{\sigma}\right) = \Phi\Big(\frac{q-\mu}{\sigma}\Big) \Rightarrow q = \mu + \sigma\Phi^{-1}(0.2) \tag{50}$$

Την τιμή της Φ και Φ^{-1} μπορούμε να την βρούμε σε κάποιον πίνακα ή χρησιμοποιώντας την Python.

4.1.1 Δύο συνεχείς τυχαίες μεταβλητές

Ορισμός 4.2 (από κοινού συνάρτηση πυκνότητας πιθανότητας)**.** *Έστω X, Y δύο τυχαίες μεταβλητές με συνάρτηση πυκνότητας πιθανότητας, f_x και f_y, αντίστοιχα. Η συνάρτηση πυκνότητας πιθανότητας για το ζευγάρι (X, Y) ορίζουμε την* από κοινού συνάρτηση πυκνότητας πιθανότητας *$f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ η οποία δίνει την πιθανότητα*

$$\mathbb{P}_{X,Y}\left(X \in A, Y \in B\right) = \int_A \int_B f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y\,. \tag{51}$$

Για την από κοινού συνάρτηση πυκνότητας πιθανότητας *πρέπει να ισχύει $f_{X,Y}(x, y) \geq 0$ για κάθε $x, y \in \mathbb{R}$ και*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1\,. \tag{52}$$

Ορισμός 4.3 (ανεξαρτησία)**.** *Οι X, Y είναι* ανεξάρτητες *αν και μόνο αν $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, για κάθε $x, y \in \mathbb{R}$.*

Ορισμός 4.4 (οριακή συνάρτηση κατανομής)**.** *Αν γνωρίζουμε την* από κοινού συνάρτηση πυκνότητας πιθανότητας *μπορούμε να βρούμε την f_X από την σχέση,*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}y\,, \tag{53}$$

και αντίστοιχα την f_Y. Οι f_X και f_Y λέγονται οριακές συναρτήσεις κατανομής *(marginal distribution functions).*

4.2 Μέση τιμή και διασπορά τυχαίας μεταβλητής

Ορισμός 4.5 (μέση τιμή)**.** *Η μέση τιμή μιας συνεχούς τυχαίας μεταβλητής με συνάρτηση πυκνότητας πιθανότητας f_X ορίζεται ως,*

$$\mu = \mathbb{E}_{f_X}\left[X\right] = \int x f_X(x)\, \mathrm{d}x\,. \tag{54}$$

Ορισμός 4.6 (διακύμανση)**.** *Η διακύμανση μιας συνεχούς τυχαίας μεταβλητής με συνάρτηση πυκνότητας πιθανότητας f_X ορίζεται ως,*

$$\sigma^2 = \mathbb{V}_{f_X}\left[X\right] = \mathbb{E}_{f_X}\left[(X - \mu)^2\right] = \int (x - \mu)^2 f_X(x)\, \mathrm{d}x\,. \tag{55}$$

Η τυπική απόκλιση της X ορίζεται ως $\sigma = \sqrt{\sigma^2}$.

Θεώρημα 4.2. *Για την τυχαία μεταβλητή $Y = g(X)$, η μέση τιμή της δίνεται από την σχέση,*

$$\mathbb{E}_{f_X}\left[Y\right] = \mathbb{E}_{f_X}\left[g(X)\right] = \int g(x) f_X(x)\, \mathrm{d}x\,. \tag{56}$$

Για την διακύμανση μπορούμε να δείξουμε ότι,

$$\mathbb{V}_{f_X}\left[X\right] = \mathbb{E}_{f_X}\left[X^2\right] - \mu^2\,, \tag{57}$$

και για $a, b \in \mathbb{R}$

$$\mathbb{V}_{f_X}\left[aX + b\right] = a^2 \mathbb{V}_{f_X}\left[X\right]\,. \tag{58}$$

4.2.1 Μέση τιμή και διασπορά δείγματος

Έστω τυχαίες μεταβλητές $X_1, X_2, \ldots, X_N$, ανεξάρτητες ανά δύο και $X_i \sim F_X, i = 1, \ldots, N$. Η μέση τιμή των X_i είναι $\mathbb{E}\left[X_i\right] = \mu$ και η διακύμανση $\mathbb{V}\left[X_i\right] = \sigma^2$. Γράφουμε ότι οι X_i είναι *iid* (independent, identically distributed).

Ορισμός 4.7. *(μέση τιμή και διακύμανση δείγματος) Ορίζουμε δύο νέες τυχαίες μεταβλητές, την* μέση τιμή του δείγματος,

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i\,, \tag{59}$$

και την διακύμανση του δείγματος,

$$S_N^2 = \frac{1}{N - 1} \sum_{i=1}^{N} (X_i - \bar{X}_N)^2\,. \tag{60}$$

Πρόταση 4.1. *Για τις τυχαίες μεταβλητές $\bar{X}_N$ και S_N^2 ισχύουν οι σχέσεις,*

$$\mathbb{E}\left[\bar{X}_N\right] = \mu, \quad \mathbb{V}\left[\bar{X}_N\right] = \frac{\sigma^2}{N}, \quad \mathbb{E}\left[S_N^2\right] = \sigma^2\,. \tag{61}$$

4.3 Νόμος των μεγάλων αριθμών

Θεώρημα 4.3 (ασθενής νόμος των μεγάλων αριθμών). *Η τυχαία μεταβλητή $\bar{X}_N$ συγκλίνει στην τιμή μ με την έννοια,*

$$\lim_{N\to\infty} \mathbb{P}\left(|\bar{X}_N - \mu| < \varepsilon\right) = 1\,, \quad \forall \varepsilon > 0\,. \tag{62}$$

Θεώρημα 4.4 (ισχυρός νόμος των μεγάλων αριθμών). *Η τυχαία μεταβλητή $\bar{X}_N$ συγκλίνει στην τιμή μ με την έννοια,*

$$\mathbb{P}\left(\lim_{N\to\infty} \bar{X}_N = \mu\right) = 1\,. \tag{63}$$

4.4 Κεντρικό οριακό θεώρημα

Το κεντρικό οριακό θεώρημα μας δίνει την κατανομή της $\bar{X}_N$ για μεγάλο N. Συγκεκριμένα,

$$\bar{X}_N \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})\,. \tag{64}$$

Θεώρημα 4.5. *Η κατανομή της τυχαίας μεταβλητή $\sqrt{N}(\bar{X}_N - \mu)$ συγκλίνει στην κανονική κατανομή $\mathcal{N}(0,\sigma^2)$ με την έννοια,*

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{\bar{X}_N - \mu}{\frac{\sigma}{\sqrt{N}}} \le z\right) = \Phi\left(z\right)\,. \tag{65}$$

5 Εβδομάδα 05/13

5.1 Κανονική κατανομή

Για την κανονική κατανομή ισχύει ότι το 68.3% του πληθυσμού βρίσκεται στο διάστημα $(\mu - \sigma, \mu + \sigma)$, το 95.4% στο διάστημα $(\mu - 2\sigma, \mu + 2\sigma)$, και το 99.7% στο διάστημα $(\mu - 3\sigma, \mu + 3\sigma)$.

5.1.1 Άθροισμα τυχαίων μεταβλητών που ακολουθούν την κανονική κατανομή

Θεώρημα 5.1. *Η τυχαία μεταβλητή $X + Y$ με $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ και $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ ακολουθεί την κατανομή $\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.*

5.2 Συσχέτιση

Ορισμός 5.1 (Συνδιακύμανση). *Η συνδιακύμανση δύο τυχαίων μεταβλητών X, Y ορίζεται ως*

$$\mathrm{Cov}\left[X,Y\right] = \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] = \mathbb{E}\left[XY\right] - \mu_X\mu_Y\,. \tag{66}$$

Ορισμός 5.2 (Pearson correlation coefficient). *Ο συντελεστής συσχέτισης ορίζεται ως*

$$\rho_{X,Y} = \frac{\mathrm{Cov}\left[X,Y\right]}{\sigma_X\sigma_Y}\,. \tag{67}$$

Άσκηση 5.1. *Για τις τυχαίες μεταβλητές $S = \sum_{i=1}^N a_i X_i$ και $T = \sum_{i=1}^M b_i Y_i$ δείξτε ότι*

$$\mathrm{Cov}\left[S,T\right] = \sum_{i=1}^N \sum_{j=1}^M a_i b_j \,\mathrm{Cov}\left[X_i, Y_j\right]\,, \tag{68}$$

και

$$\mathbb{V}\left[S\right] = \sum_{i=1}^N a_i^2 \mathbb{V}\left[X_i\right] + 2\sum_{i=1}^N \sum_{j=1}^{i-1} a_i a_j \,\mathrm{Cov}\left[X_i, Y_j\right]\,. \tag{69}$$

Ορισμός 5.3 (Συνδιακύμανση και συντελεστής συσχέτισης δείγματος). *Για ένα δείγμα $\{(x_i, y_i)\}_{i=1}^N$, ορίζουμε την συνδιακύμανση του δείγματος ως,*

$$S_{X,Y} = \frac{1}{N-1}\sum_{i=1}^N \left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)\,, \tag{70}$$

και τον συντελεστή συσχέτισης ως,

$$r_{x,y} = \frac{S_{X,Y}}{S_X S_Y}\,, \tag{71}$$

όπου S_X και S_Y η διασπορά του δείγματος για την μεταβλητή X και Y αντίστοιχα.

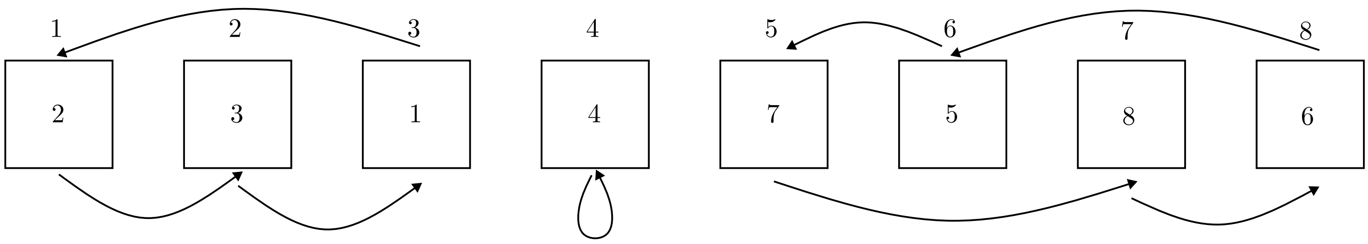
5.3 Το πρόβλημα των 100 κρατούμενων

Ο διευθυντής μιας φυλακής με 100 κρατούμενους τους δίνει το παρακάτω πρόβλημα. Σε ένα δωμάτιο βάζει 100 κλειστά κουτιά με τους αριθμούς 1 έως 100. Σε κάθε έναν κρατούμενο αναθέτει έναν αριθμό από το 1 έως το 100. Ένας ένας οι κρατούμενοι μπαίνουν στο δωμάτιο και ανοίγουν μέχρι 50 κουτιά. Αν βρουν το νούμερο τους σημειώνουν μια νίκη, αλλιώς μια ήττα. Οι κρατούμενοι βγαίνουν από διαφορετική πόρτα και δεν μπορούν να επικοινωνήσουν αφού μπουν στο δωμάτιο. Αν σημειωθούν 100 νίκες ο διευθυντής τους ελευθερώνει, αλλιώς βάζει σε όλους ισόβια.

Αρχικά θα υπολογίσουμε την πιθανότητα να σημειωθούν 100 νίκες με την υπόθεση ότι όλοι οι κρατούμενοι ανοίγουν 50 κουτιά στην τύχη. Ο κάθε κρατούμενος έχει πιθανότητα 0.5 να βρει το νούμερο του. Εφόσον οι κρατούμενοι δεν έχουν επικοινωνία μεταξύ τους, τα ενδεχόμενα είναι ανεξάρτητα. Η πιθανότητα να σημειωθούν 100 νίκες είναι ίση με το γινόμενο της πιθανότητας της κάθε νίκης. Επομένως, η πιθανότητα να ελευθερωθούν και οι 100 είναι ίση με $2^{-100} \approx 10^{-30}$.

Θα δείξουμε ότι υπάρχει μία στρατηγική η οποία μπορεί να ανεβάσει την πιθανότητα να ελευθερωθούν όλοι από 10^{-30} σε περίπου 0.3. Ο κάθε κρατούμενος ανοίγει το κουτί με νούμερο του. Στη συνέχεια ανοίγει το κουτί με το νούμερο του προηγούμενου κουτιού. Συνεχίζει με αυτό τον τρόπο μέχρι να ανοίξει το πολύ 50 κουτιά. Προφανώς, για να σημειώσει μια νίκη ο κρατούμενος, θα πρέπει να υπάρχει ένα κουτί στην αλληλουχία που να περιέχει το νούμερο του. Παρατηρήστε ότι με τον τρόπο αυτό κλείνει ένας κύκλος. Ξεκίνησε από το κουτί με το νούμερο του και κατέληξε πάλι στο ίδιο κουτί.

Στο Σχήμα 7 βλέπουμε μια περίπτωση όπου όλοι οι κύκλοι έχουν μήκος μικρότερο η ίσο του 4 για το πρόβλημα με 8 κρατούμενους αντί για 100. Μας ενδιαφέρει να μετρήσουμε πόσες αναδιατάξεις αριθμών από το 1 έως το 100 έχουν κύκλους με μέγιστο μήκος μικρότερο ή ίσο του 50.



Σχήμα 7: Το πρόβλημα των κρατούμενων για 8 κρατούμενους. Μία διάταξη των αριθμών 1 έως 8 όπου οι κρατούμενοι κερδίζουν. Ο κρατούμενος με τον αριθμό 1 ανοίγει το κουτί 1, μετά το κουτί 2, μετά το κουτί 3, και σταματάει γιατί βρήκε το νούμερο του. Ο κρατούμενος με τον αριθμό 8 ανοίγει το κουτί 8, μετά το κουτί 6, μετά το κουτί 5, μετά το κουτί 5 , και σταματάει γιατί βρήκε το νούμερο του. Όλοι οι κύκλοι έχουν μήκος μικρότερο ή ίσο του 4.

Έστω A το σύνολο όλων των αναδιατάξεων των αριθμών 1 έως 100 με κύκλους μέγιστου μήκους 50. Η πιθανότητα του συνόλου είναι ίση με $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$. Θα μετρήσουμε το μέγεθος του A^c. Έστω $\ell = 51, 52, \ldots, 100$ και A_ℓ το σύνολο που περιέχει όλες τις αναδιατάξεις των αριθμών από το 1 έως το 100 με μέγιστο μήκος κύκλου ίσο με ℓ. Στις ℓ πρώτες θέσεις βάζουμε τους αριθμούς που θα σχηματίσουν τον κύκλο μήκους ℓ. Επιλέγουμε αυτούς τους αριθμούς με $\binom{100}{\ell}$ τρόπους. Οι αριθμοί αυτοί μπορούνε να διαταχθούν με $(\ell-1)!$ τρόπους. Προσέξτε ότι είναι $(\ell-1)!$ οι τρόποι και όχι $\ell!$ γιατί οι διατάξεις είναι κυκλικές. Δείτε για παράδειγμα ότι η διάταξη 1234 είναι ίδια με την 2341. Στις $N - \ell$ θέσεις που απομένουν τοποθετούμε τους αριθμούς με $(100-\ell)!$ τρόπους. Τέλος, παρατηρούμε ότι τα A_ℓ^c είναι ξένα μεταξύ τους για $\ell > 50$ και επομένως

$$\begin{aligned}
\mathbb{P}\left(A\right) &= 1 - \mathbb{P}\left(A^c\right) \\
&= 1 - \mathbb{P}\left(\bigcup_{\ell=51}^{100} A_\ell\right) \\
&= 1 - \sum_{\ell=51}^{100} \mathbb{P}\left(A_\ell\right) \\
&= 1 - \sum_{\ell=51}^{100} \frac{1}{100!}\binom{100}{\ell}(\ell-1)!(100-\ell)! \\
&= 1 - \sum_{\ell=51}^{100} \frac{1}{\ell} \\
&= 1 - H_{100} - H_{50}\,,
\end{aligned} \tag{72}$$

όπου $H_n = \sum_{i=1}^{n} \frac{1}{i}$. Για 100 κρατούμενους υπολογίζουμε ότι η πιθανότητα είναι περίπου 0.31183.

Για N κρατούμενους, έχουμε ότι

$$\mathbb{P}\left(A\right) = 1 - H_N - H_{\frac{N}{2}}\,. \tag{73}$$

Γνωρίζουμε ότι

$$\lim_{N\to\infty}\left(H_N \log N\right) = \gamma\,, \tag{74}$$

όπου γ η σταθερά Euler. Επομένως, καθώς ο αριθμός των κρατουμένων πηγαίνει στο άπειρο,

$$\mathbb{P}\left(A\right) = 1 - \lim_{N\to\infty}\left(H_N - \log N - (H_{\frac{N}{2}} - \log\frac{N}{2}) - \log 2\right) = 1 - \log 2 \approx 0.307\,. \tag{75}$$

6 Εβδομάδα 06/13

6.1 Δεσμευμένη πιθανότητα

Θεωρούμε το πείραμα όπου ρίχνουμε δύο ζάρια, $\Omega = (\omega_1, \omega_2)$, και τις τυχαίες μεταβλητές $X(\omega) = \omega_1$ και $Y(\omega) = \omega_1 + \omega_2$. Θέλουμε να υπολογίσουμε την πιθανότητα η X να είναι μικρότερη ή ίση του 3, γνωρίζοντας ότι η Y είναι ίση με 6. Αυτή είναι μια δεσμευμένη πιθανότητα στο γεγονός $Y = 3$ και γραφουμε $\mathbb{P}\big(X(\omega) \leq 3 \mid Y(\omega) = 6\big)$. Το γεγονός $Y = 6$ περιορίζει τον δειγματικό χώρο στα ενδεχόμενα $B = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$. Από αυτά, τα $A = \{(1,5),(2,4),(3,3)\}$. Επομένως, η πιθανότητα είναι $\frac{|A|}{|B|}$.

Ορισμός 6.1 (δεσμευμένη συνάρτηση μάζας πιθανότητας)**.** *Η δεσμευμένη συνάρτηση μάζας πιθανότητας ορίζεται ως*

$$f_{X\,|\,Y}(x\,|\,y) = \mathbb{P}_{X\,|\,Y}\big(X = x\,|\,Y = y\big) = \frac{\mathbb{P}_{X,Y}\left(X = x, Y = y\right)}{\mathbb{P}_{Y}\left(Y = y\right)} \tag{76}$$

$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
1	✖	✖	✖	✖	✖	✖
2	✖	✖	✖	✖	✖	✖
3	✖	✖	✖	✖	✖	✖
4	✖	✖	✖	✖	✖	✖
5	✖	✖	✖	✖	✖	✖
6	✖	✖	✖	✖	✖	✖

(α′) $\mathbb{P}\left(X(\omega) = 1\right) = \frac{6}{36}$

$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
1	✖	✖	✖	✖	✖	✖
2	✖	✖	✖	✖	✖	✖
3	✖	✖	✖	✖	✖	✖
4	✖	✖	✖	✖	✖	✖
5	✖	✖	✖	✖	✖	✖
6	✖	✖	✖	✖	✖	✖

(β′) $\mathbb{P}\left(Y(\omega) \geq 6\right) = \frac{26}{36}$

$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
1	✖	✖	✖	✖	✖	✖
2	✖	✖	✖	✖	✖	✖
3	✖	✖	✖	✖	✖	✖
4	✖	✖	✖	✖	✖	✖
5	✖	✖	✖	✖	✖	✖
6	✖	✖	✖	✖	✖	✖

(γ′) $\mathbb{P}\left(X(\omega) \leq 3, Y(\omega) \geq 6\right) = \frac{9}{36}$

$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
1					✖	
2				✖		
3			✖			
4		✖				
5	✖					
6						

(δ′) $\mathbb{P}\left(X(\omega) \leq 3 \,|\, Y(\omega) = 6\right) = \frac{3}{5}$

$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
1	✖	✖	✖			
2	✖	✖	✖			
3	✖	✖	✖			
4	✖	✖	✖			
5	✖	✖	✖			
6	✖	✖	✖			

(ε′) $\mathbb{P}\left(Y(\omega) = 6 \,|\, X(\omega) \leq 3\right) = \frac{3}{18}$

Πίνακας 5: Τα κελιά με το σύμβολο ✖ αποτελούν το δειγματικό χώρο και τα κόκκινα κελιά τα ενδεχόμενα που μας ενδιαφέρουν.

Ορισμός 6.2 (δεσμευμένη συνάρτηση πυκνότητας πιθανότας)**.** *Η δεσμευμένη συνάρτηση μάζας πιθανότητας ορίζεται ως*

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\,, \tag{77}$$

για τα $y \in \mathbb{R}$ για τα οποία $f_Y(y) \neq 0$. Για $A \subset \mathbb{R}$,

$$\mathbb{P}_{X\,|\,Y}\left(X \in A \,|\, Y = y\right) = \int_A f_{X\,|\,Y}(x\,|\,y)\,\mathrm{d}x\,. \tag{78}$$

Παράδειγμα 6.1. *Οι τυχαίες μεταβλητές X, Y έχουν απο κοινού συνάρτηση πυκνότητας πιθανότητας*

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{αλλιώς} \end{cases}\,. \tag{79}$$

Για να βρούμε την $f_{X\,|\,Y}$ υπολογίζουμε πρώτα την περιθώρια πυκνότητα πιθανότητας για $0 \leq y \leq 1$,

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,\mathrm{d}y = \int_0^1 1\,\mathrm{d}x = 1\,. \tag{80}$$

Επομένως,

$$f_Y(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{αλλιώς} \end{cases}\,. \tag{81}$$

Από τον Ορισμό 6.2, η δεσμευμένη συνάρτηση πυκνότητας ορίζεται για $x \in [0,1]$, και είναι ίση με 1.

Παράδειγμα 6.2. *Οι τυχαίες μεταβλητές X, Y έχουν απο κοινού συνάρτηση πυκνότητας πιθανότητας*

$$f_{X,Y}(x,y) = \begin{cases} x + y & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{αλλιώς} \end{cases}\,. \tag{82}$$

Θέλουμε να βρούμε την πιθανότητα $\mathbb{P}_{X|Y}\left(X < \frac{1}{4} \,|\, Y = \frac{1}{3}\right)$. Για να βρούμε την $f_{X\,|\,Y}$ υπολογίζουμε πρώτα την περιθώρια πυκνότητα πιθανότητας για $0 \leq y \leq 1$,

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,\mathrm{d}y = \int_0^1 x + y\,\mathrm{d}x = y + \frac{1}{2}\,. \tag{83}$$

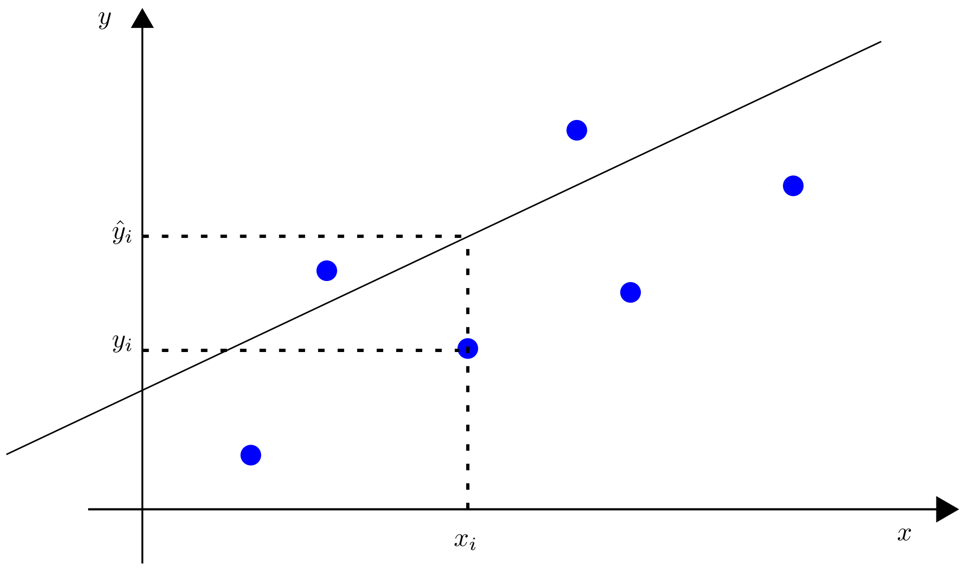
Από τον Ορισμό 6.2,

$$f_{X|Y}(x\,|\,y) = \begin{cases} \dfrac{x+y}{y+\frac{1}{2}} & 0 \le y \le 1 \\ 0 & \textit{αλλιώς} \end{cases}. \tag{84}$$

Επομένως,

$$\mathbb{P}_{X|Y}\left(X < \frac{1}{4}\,|\,Y = \frac{1}{3}\right) = \int_{-\infty}^{\frac{1}{4}} f_{X|Y}(x\,|\,\frac{1}{3})\,\mathrm{d}x = \int_{0}^{\frac{1}{4}} \frac{x+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}}\,\mathrm{d}x = \dots \tag{85}$$

7 Εβδομάδα 07/13



Σχήμα 8: Ένα σύνολο σημείων $\{(x_i, y_i)\}_{i=1}^n$ και $\hat{y}_i$ η πρόβλεψη του μοντέλου στο σημείο x_i.

7.1 Απλή γραμμική παλινδρόμηση

Έστω έναν σύνολο δεδομένων $\{(x_i, y_i)\}_{i=1}^n$ με $x_i, y_i \in \mathbb{R}$. Υποθέτουμε ότι τα δεδομένα προέρχονται απο παρατηρήσεις τυχαίων μεταβλητών (X_i, Y_i) με απο κοινού συνάρτηση πυκνότητας πιθανότητας $p_{X,Y}$ η οποία μας είναι άγνωστη. Η μεταβλητή X ονομάζεται *επεξηγηματική* μεταβλητή και η Y μεταβλητή *απόκρισης*. Αποφεύγουμε να χρησιμοποιήσουμε τους όρους ανεξάρτητη και εξαρτημένη μεταβλητή για τις X και Y αντίστοιχα, ώστε να αποφύγουμε την σύγχυση με τον ορισμό της ανεξαρτησίας για τυχαίες μεταβλητές. Ιδανικά θα θέλαμε να μάθουμε την συνάρτηση $p_{X,Y}$ βασισμένοι στο δοσμένο σύνολο δεδομένων. Θα ξεκινήσουμε από ένα πιο εύκολο πρόβλημα και θα προσπαθήσουμε να μάθουμε την συνάρτηση,

$$r(x) := \mathbb{E}\left[Y\,|\,X = x\right]. \tag{86}$$

Υποθέτουμε ότι οι X και Y συνδέονται από την σχέση

$$Y = \beta_0 + \beta_1 X + E, \tag{87}$$

όπου η E είναι τυχαία μεταβλητή με $\mathbb{E}[E] = 0$ και $\mathbb{V}[E] = \sigma^2$. Υποθέτουμε δηλαδή ότι η r είναι γραμμική στο x, ότι δηλαδή υπάρχουν β_0 και β_1 έτσι ώστε

$$r(x; \beta_0, \beta_1) = \beta_0 + \beta_1 x. \tag{88}$$

Ο σκοπός είναι να βρούμε $\hat{\beta}_0$ και $\hat{\beta}_1$ ώστε να εκτιμούν τις ποσότητες β_0 και β_1. Παρατηρήστε ότι οι ποσότητες $\hat{\beta}_0$ και $\hat{\beta}_1$ θα είναι τυχαίες μεταβλητές· αλλάζοντας το αρχικό σύνολο δεδομένων, παίρνουμε διαφορετικές τιμές.

Ορισμός 7.1 (πρόβλεψη)**.** *Για δοσμένα $\hat{\beta}_0$ και $\hat{\beta}_1$, ορίζουμε την πρόβλεψη του μοντέλου στο σημείο x,*

$$\hat{y} := r(x; \hat{\beta}_0, \hat{\beta}_1) = \hat{\beta}_0 + \hat{\beta}_1 x. \tag{89}$$

Ορισμός 7.2 (σφάλμα)**.** *Ορίζουμε το σφάλμα ως*

$$\mathcal{E}(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \varepsilon_i^2(\hat{\beta}_0, \hat{\beta}_1), \tag{90}$$

όπου

$$\varepsilon_i(\hat\beta_0,\hat\beta_1) := y_i - \hat y_i := y_i - r(x_i,\hat\beta_0,\hat\beta_1) = y_i - \hat\beta_0 - \hat\beta_1 x_i\,. \tag{91}$$

Στο σχήμα 8 απεικονίζονται τα δεδομένα, το γραμμικό μοντέλο, και το σφάλμα του μοντέλου στα δεδομένα.

Θεώρημα 7.1 (ελάχιστα τετράγωνα)**.** *Ορίζουμε $\hat\beta_0$ και $\hat\beta_1$ ως,*

$$\hat\beta_0,\hat\beta_1 = \operatorname*{arg\,min}_{\beta_0,\beta_1} \mathcal{E}(\beta_0,\beta_1)\,. \tag{92}$$

Το πρόβλημα ελαχιστοποίησης έχει μοναδική λύση,

$$\begin{aligned}\hat\beta_0 &= \bar Y - \frac{S_{XY}}{S_X^2}\bar X\,,\\ \hat\beta_1 &= \frac{S_{XY}}{S_X^2}\,.\end{aligned} \tag{93}$$

Επιπροσθέτως, η διακύμανση της τυχαίας μεταβλητής E μπορεί να εκτιμηθεί από την εκτιμήτρια

$$\widehat{\sigma^2} = \frac{1}{n-2}\sum_{i=1}^n \varepsilon_i^2\,. \tag{94}$$

Απόδειξη. Υπολογίζουμε τις παραγώγους της συνάρτησης E, τις θέτουμε ίσες με μηδέν, και λύνουμε ώς προς β_0 και β_1. Για την παράγωγο ως προς β_0 έχουμε,

$$\begin{aligned}\frac{\partial\mathcal{E}}{\partial\beta_0}(\beta_0,\beta_1) &= \frac{\partial}{\partial\beta_0}\sum_{i=1}^n (y_i-\beta_0+\beta_1 x_i)^2\\ &= -2\sum_{i=1}^n y_i-\beta_0-\beta_1 x_i = 0\\ \Rightarrow \beta_0 &= \bar Y - \beta_1\bar X\,.\end{aligned} \tag{95}$$

Για την παράγωγο ως προς β_1 έχουμε,

$$\begin{aligned}\frac{\partial\mathcal{E}}{\partial\beta_1}(\beta_0,\beta_1) &= \frac{\partial}{\partial\beta_0}\sum_{i=1}^n (y_i-\beta_0+\beta_1 x_i)^2\\ &= -2\sum_{i=1}^n (y_i-\beta_0-\beta_1 x_i)\,x_i = 0\\ \Rightarrow \beta_1\sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_iy_i - \beta_0\sum_{i=1}^n x_i\\ \Rightarrow \beta_1\sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_iy_i - \beta_0 n\bar X\\ \Rightarrow \beta_1\sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_iy_i - (\bar Y-\beta_1\bar X)n\bar X\\ \Rightarrow \beta_1\left(\sum_{i=1}^n x_i^2 - n\bar X^2\right) &= \sum_{i=1}^n x_iy_i - n\bar Y\bar X\\ \Rightarrow \beta_1 &= \frac{\sum_{i=1}^n x_iy_i - n\bar Y\bar X}{\sum_{i=1}^n x_i^2 - n\bar X^2}\\ \beta_1 &= \frac{S_{XY}}{S_X^2}\,.\end{aligned} \tag{96}$$

Για να δείξουμε ότι το ακρότατο είναι ελάχιστο αρκεί να δείξουμε ότι ο Εσσιανός πίνακας είναι θετικα ορισμένος.

Η απόδειξη για την εκτιμήτρια $\widehat{\sigma^2}$ δίνεται στην γενική περίπτωση στην απόδειξη του θεωρήματος 7.2. □

Άσκηση 7.1. *Δείξτε ότι ο Εσσιανός πίνακας στην απόδειξη του θεωρήματος 7.1 είναι θετικά ορισμένος.*

Ορισμός 7.3 (αθροίσματα τετραγώνων). *Ορίζουμε τα συνολικό άθροισμα τετραγώνων,*

$$SS_{tot} \coloneqq \sum_{i=1}^{n} \left(y_i - \bar{Y}\right)^2 = \|\boldsymbol{y} - \bar{Y}\boldsymbol{e}_n\|_2^2 \,, \tag{97}$$

το άθροισμα τετραγώνων των υπολοίπων,

$$SS_{res} \coloneqq \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 = \mathcal{E} \,, \tag{98}$$

το άθροισμα τετραγώνων της παλινδρόμησης,

$$SS_{reg} \coloneqq \sum_{i=1}^{n} \left(\hat{y}_i - \bar{Y}\right)^2 = \|\hat{\boldsymbol{y}} - \bar{Y}\boldsymbol{e}_n\|_2^2 \,, \tag{99}$$

όπου $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$, $\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^\top$, και $\boldsymbol{e}_n = (1, 1, \ldots, 1) \in \mathbb{R}^n$.

Παρατήρηση 7.1. *Παρατηρήστε ότι το SS_{tot} δίνει το σφάλμα αν θεωρήσουμε ότι το μοντέλο της δεσμευμένης μέσης τιμής είναι σταθερό και ίσο με $\bar{Y}$.*

Πρόταση 7.1. *Για το συνολικό άθροισμα τετραγώνων ισχύει ότι,*

$$SS_{tot} = SS_{res} + SS_{reg} \,. \tag{100}$$

Απόδειξη. Το συνολικό άθροισμα τετραγώνων μπορεί να γραφτεί ως

$$\begin{aligned} SS_{tot} &= \sum_{i=1}^{n} \left(y_i - \hat{y}_i + \hat{y}_i - \bar{Y}\right)^2 \\ &= SS_{res} + SS_{reg} + 2\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{Y}) \,. \end{aligned} \tag{101}$$

Θα δείξουμε ότι ο τρίτος όρος είναι ίσος με το μηδέν. Παρατηρούμε ότι,

$$\hat{y}_i - \bar{Y} = \hat{\beta}_1(x_i - \bar{X}) \,, \tag{102}$$

αφαιρώντας κατά μέλη τις σχέσεις $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ και $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$, όπου η δεύτερη σχέση ισχύει από την πρώτη αν αθροίσουμε για $i = 1, \ldots, n$ και διαιρέσουμε με n. Επίσης, ισχύει ότι,

$$y_i - \hat{y}_i = y_i - \bar{Y} - (\hat{y}_i - \bar{Y}) = y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{X}) \,. \tag{103}$$

Αντικαθιστώντας τις σχέσεις (102) και (103) στον τρίτο όρο της τελευταίας ισότητας της σχέσης (101), έχουμε,

$$\begin{aligned} \sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{Y}) &= \sum_{i=1}^{n} \hat{\beta}_1(x_i - \bar{X}) \left[(y_i - \bar{Y}) - \hat{\beta}_1(x_i - \bar{X})\right] \\ &= \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y}) - \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{X})^2 \\ &= \hat{\beta}_1 S_{XY} - \hat{\beta}_1^2 S_X^2 \\ &= 0 \,, \end{aligned} \tag{104}$$

όπου η τελευταία σχέση ισχύει από το θεώρημα 7.1. □

Ορισμός 7.4 (συντελεστής προσιορισμού). *Ορίζουμε τον συντελεστή προσδιορισμού ως*

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \,. \tag{105}$$

Η τιμή R^2 δείχνει πόσο διαφέρει το σφάλμα του γραμμικού μοντέλου σε σχέση με το σφάλμα του σταθερού μοντέλου. Αν τα δύο σφάλματα είναι κοντά, τότε το γραμμικό μοντέλο δεν προσφέρει περισσότερη πληροφορία από το σταθερό μοντέλο και το R^2 είναι κοντά στο 0. Στην αντίθετη περίπτωση, όπου το σφάλμα του γραμμικού μοντέλου είναι κοντά στο μηδέν ενώ του σταθερού απέχει από το μηδέν, τότε το γραμμικό μοντέλο προσφέρει περισσότερη πληροφορία και το R^2 είναι κοντά στο 1.

7.2 Πολλαπλή γραμμική παλινδρόμηση

Σε αντίθεση με την απλή γραμμική παλινδρόμηση όπου η επεξηγηματική μεταβλητή είναι μία, σε αυτή την ενότητα θα θεωρήσουμε την περίπτωση πολλών επεξηγηματικών μεταβλητών. Έστω έναν σύνολο δεδομένων $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ με $\boldsymbol{x}_i \in \mathbb{R}^m$ και $y_i \in \mathbb{R}$. Τα διανύσματα $\boldsymbol{x}_i$ είναι πραγματοποιήσεις των τυχαίων μεταβλητών $\boldsymbol{X}_i$ με κοινή κατανομή p_X. Όμοια, οι τιμές y_i είναι πραγματοποιήσεις των τυχαίων μεταβλητών Y_i με κοινή κατανομή p_Y.

Επεκτείνουμε το διάνυσμα $\boldsymbol{X}$ θέτοντας την πρώτη θέση ίση με 1, δηλαδή $(x_1, \ldots, x_m) \mapsto (1, x_1, \ldots, x_m)$. Με τον τρόπο αυτό μπορούμε να γράψουμε το μοντέλο της σχέσης (87) ως,

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m + E = \boldsymbol{X}^\top \boldsymbol{\beta} + E\,, \tag{106}$$

όπου $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)^\top$ και E τυχαία μεταβλητή με $\mathbb{E}[E] = \boldsymbol{0}$ και $\mathbb{V}[E] = \sigma^2 I_{m+1}$. Υπολογίζουμε την πρόβλεψη του μοντέλου για την παράμετρο $\hat{\boldsymbol{\beta}}$ για κάθε $\boldsymbol{x}_i$,

$$\hat{y}_i = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}\,, \tag{107}$$

και γράφουμε,

$$\hat{\boldsymbol{y}} = A\hat{\boldsymbol{\beta}}, \tag{108}$$

όπου

$$A = \begin{bmatrix} \boldsymbol{x_1}^\top \\ \vdots \\ \boldsymbol{x_n}^\top \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \ldots & x_{1,m} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & \ldots & x_{n,m} \end{bmatrix}. \tag{109}$$

Ορισμός 7.5 (σφάλμα)**.** *Ορίζουμε το σφάλμα ως*

$$\mathcal{E}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n \varepsilon_i^2(\hat{\boldsymbol{\beta}})\,, \tag{110}$$

όπου

$$\varepsilon_i(\hat{\boldsymbol{\beta}}) \coloneqq y_i - \hat{y}_i \coloneqq y_i - r(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) = y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}\,. \tag{111}$$

Παρατήρηση 7.2. *Παρατηρήστε ότι το σφάλμα μπορεί να γραφτεί ως*

$$\mathcal{E}(\hat{\boldsymbol{\beta}}) = \|\boldsymbol{y} - A\hat{\boldsymbol{\beta}}\|_2^2\,. \tag{112}$$

Θεώρημα 7.2 (ελάχιστα τετράγωνα)**.** *Για $n \geq m+1$ ορίζουμε το διάνυσμα των βέλτιστων παραμέτρων $\hat{\boldsymbol{\beta}}$ ως,*

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \mathcal{E}(\boldsymbol{\beta})\,. \tag{113}$$

Υποθέτουμε ότι ο A έχει γραμμικά ανεξάρτητες στήλες, δηλαδή $\operatorname{rank}(A) = m+1$. Το πρόβλημα ελαχιστοποίησης έχει μοναδική λύση,

$$\hat{\boldsymbol{\beta}} = (A^\top A)^{-1} A^\top \boldsymbol{y}\,. \tag{114}$$

Επιπροσθέτως, η διακύμανση της τυχαίας μεταβλητής E μπορεί να εκτιμηθεί από την εκτιμήτρια

$$\widehat{\sigma^2} = \frac{1}{n-m-1} \sum_{i=1}^n \varepsilon_i^2\,. \tag{115}$$

Απόδειξη. Υπολογίζουμε τα ακρότατα θέτοντας την πρώτη παράγωγο της $\mathcal{E}$,

$$\begin{aligned} D_{\boldsymbol{\beta}}\mathcal{E}(\boldsymbol{\beta}) &= -2\boldsymbol{y}^\top A + 2\boldsymbol{\beta}^\top A^\top A = 0 \\ \Rightarrow A^\top A\boldsymbol{\beta} &= A\boldsymbol{y} \\ \Rightarrow \hat{\boldsymbol{\beta}} &= (A^\top A)^{-1} A^\top \boldsymbol{y}\,. \end{aligned} \tag{116}$$

Ο Εσσιανός πίνακας,

$$D^2_{\boldsymbol{\beta}}\mathcal{E}(\boldsymbol{\beta}) = 2A^\top A\,, \tag{117}$$

είναι θετικά ορισμένος. Επομένως το ακρότατο είναι ελάχιστο.

Παρατηρούμε ότι ο πίνακας $P = A(A^\top A)^{-1}A^\top$ είναι πίνακας προβολής, δηλαδή $P^2 = P$. Το διάνυσμα των σφαλμάτων γράφεται ως,

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - A\hat{\boldsymbol{\beta}} = \boldsymbol{y} - A(A^\top A)^{-1}A^\top \boldsymbol{y} = (I_n - P)\,\boldsymbol{y}\,. \tag{118}$$

Για το διάνυσμα $\boldsymbol{\varepsilon}$ ισχύει ότι,

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\varepsilon} \,|\, A\right] &= \mathbb{E}\left[\boldsymbol{y} - A\hat{\boldsymbol{\beta}} \,|\, A\right] \\
&= \mathbb{E}\left[\boldsymbol{y} \,|\, A\right] - \mathbb{E}\left[A(A^\top A)^{-1}A^\top \boldsymbol{y} \,|\, A\right] \\
&= A\boldsymbol{\beta} - A(A^\top A)^{-1}A^\top \mathbb{E}\left[\boldsymbol{y} \,|\, A\right] \\
&= A\boldsymbol{\beta} - A(A^\top A)^{-1}A^\top A\boldsymbol{\beta} = 0\,.
\end{aligned} \tag{119}$$

Ο πίνακας $I_n - P$ είναι συμμετρικός και είναι επίσης πίνακας προβολής. Ισχύει ότι,

$$\begin{aligned}
\mathrm{Cov}\left[\boldsymbol{\varepsilon} \,|\, A\right] &= \mathrm{Cov}\left[(I_n - P)\,\boldsymbol{y} \,|\, A\right] \\
&= (I_n - P)\,\mathrm{Cov}\left[\boldsymbol{y} \,|\, A\right](I_n - P) \\
&= \sigma^2\,(I_n - P)\;.
\end{aligned} \tag{120}$$

Παίρνοντας το trace των πινάκων στην παραπάνω σχέση,

$$\begin{aligned}
\mathrm{tr}\left(\mathrm{Cov}\left[\boldsymbol{\varepsilon} \,|\, A\right]\right) &= \sigma^2\,\mathrm{tr}\,(I_n - P) \Rightarrow \\
\sigma^2 &= \frac{\sum_{i=1}^n \mathbb{V}\left[\varepsilon_i \,|\, A\right]}{\mathrm{tr}\,(I_n - P)} \\
&= \frac{\sum_{i=1}^n \mathbb{E}\left[\varepsilon_i^2 \,|\, A\right]}{\mathrm{tr}\,(I_n - P)} \\
&= \mathbb{E}\left[\frac{\sum_{i=1}^n \varepsilon_i^2}{\mathrm{tr}\,(I_n - P)} \,|\, A\right] \\
&= \mathbb{E}\left[\frac{1}{n-m-1}\sum_{i=1}^n \varepsilon_i^2 \,|\, A\right]\,,
\end{aligned} \tag{121}$$

διότι για το trace ισχύει ότι

$$\begin{aligned}
\mathrm{tr}\,(I_n - P) &= \mathrm{tr}\,(I_n) - \mathrm{tr}\,(P) \\
&= n - \mathrm{tr}\left(A(A^\top A)^{-1}A^\top\right) \\
&= n - \mathrm{tr}\left(A^\top A(A^\top A)^{-1}\right) \\
&= n - \mathrm{tr}\,(I_{m+1}) \\
&= n - m - 1\,.
\end{aligned} \tag{122}$$

Καταλήγουμε ότι μία εκτιμήτρια του σ^2 είναι,

$$\widehat{\sigma^2} = \frac{1}{n-m-1}\sum_{i=1}^n \varepsilon_i^2\,. \tag{123}$$

□

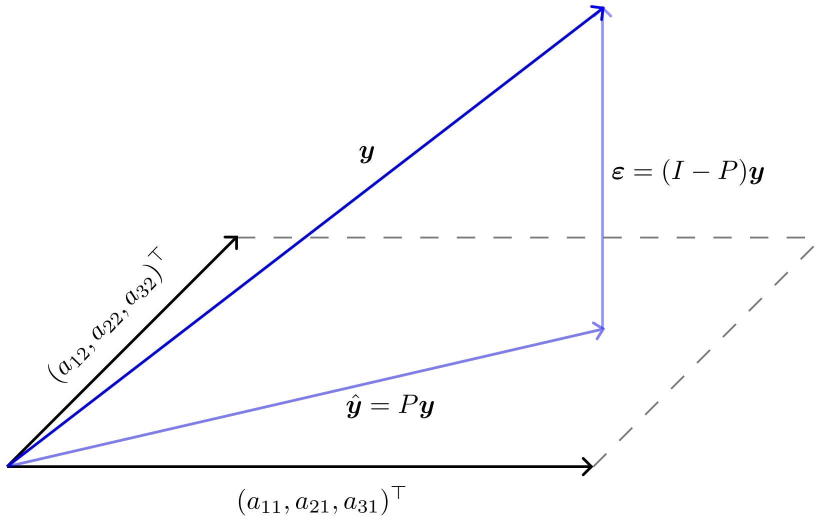
Ερώτηση 7.1. *Δείξτε ότι ο πίνακας* $A^\top A$ *είναι θετικά ορισμένος για κάθε* $A \in \mathbb{R}^{n,m}$.

Πρόταση 7.2. *Για το διάνυσμα των σφαλμάτων* $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$ *ισχύει ότι,*

$$A^\top \boldsymbol{\varepsilon} = 0\,. \tag{124}$$

Απόδειξη. Παρατηρούμε ότι

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - A\hat{\boldsymbol{\beta}}\,. \tag{125}$$



Σχήμα 9: Το διάνυσμα των δεδομένων $\boldsymbol{y}$ είναι το άθροισμα δύο διανυσμάτων, του διανύσματος των προβλέψεων $\hat{\boldsymbol{y}}$, που είναι γραμμικός συνδυασμός των στηλών του πίνακα A, και του διανύσματος των σφαλμάτων $\boldsymbol{\varepsilon}$, που είναι κάθετο στο $\hat{\boldsymbol{y}}$. Το διάνυσμα $\hat{\boldsymbol{y}}$ είναι η προβολή του $\boldsymbol{y}$ στον χώρο στηλών μέσω του πίνακα προβολής P, και το διάνυσμα $\boldsymbol{\varepsilon}$ είναι η προβολή του $\boldsymbol{y}$ στον μηδενόχωρο του χώρου στηλών μέσω του πίνακα προβολής $I - P$.

Εφόσον, $\hat{\boldsymbol{\beta}} = (A^\top A)^{-1} A^\top \boldsymbol{y}$,

$$\begin{aligned} A^\top A\hat{\boldsymbol{\beta}} &= A^\top \boldsymbol{y} \Rightarrow \\ A^\top(\boldsymbol{y} - A\hat{\boldsymbol{\beta}}) &= 0 \Rightarrow \\ A^\top \boldsymbol{\varepsilon} &= 0\,. \end{aligned} \tag{126}$$

□

Η Πρόταση 7.2 λέει ότι το διάνυσμα του σφάλματος ανήκει στον αριστερό μηδενόχωρο ή αλλιώς στον μηδενόχωρο του χώρου στηλών. Κάθε διάνυσμα που γράφεται σαν γραμμικός συνδυασμός των στηλών του A, δηλαδή κάθε διάνυσμα $\hat{\boldsymbol{y}} = A\hat{\boldsymbol{\beta}}$, είναι κάθετο στο διάνυσμα του σφάλματος, $\hat{\boldsymbol{y}} \perp \boldsymbol{\varepsilon}$.

8 Εβδομάδα 09/13

Πρόοδος

9 Εβδομάδα 09/13

9.1 Γενικευμένη γραμμική παλινδρόμηση

Μπορούμε να γενικεύσουμε την γραμμική παλινδρόμηση θεωρώντας το μοντέλο

$$r(\boldsymbol{x}) = g^{-1}(\boldsymbol{x}^\top \boldsymbol{\beta})\,, \tag{127}$$

για οποιαδήποτε αντιστρέψιμη συνάρτηση $g : \mathbb{R} \to \mathbb{R}$. Θεωρούμε το γραμμικό μοντέλο $\hat{\boldsymbol{\eta}} = \boldsymbol{x}^\top \boldsymbol{\beta}$ και βρίσκουμε τις βέλτιστες παραμέτρους για τα δεδομένα $\boldsymbol{\eta} = g(\boldsymbol{y})$.

9.2 Λογιστική παλινδρόμηση

Θεωρούμε το σύνολο δεδομένων $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ με $\boldsymbol{x}_i \in \mathbb{R}^m$ και $y_i \in \{0, 1\}$. Υποθέτουμε ότι τα δεδομένα προέρχονται από τις τυχαίες μεταβλητές $(\boldsymbol{X}_i, Y_i)$ με δεσμευμένη συνάρτηση μάζας πιθανότητας

$$\begin{aligned} p(y \,|\, \boldsymbol{x}) &:= \mathbb{P}\left(Y = y \,|\, \boldsymbol{X} = \boldsymbol{x}\right) \\ &= \Big(\mathbb{P}\left(Y = 0 \,|\, \boldsymbol{X} = \boldsymbol{x}\right)\Big)^{1-y} + \Big(\mathbb{P}\left(Y = 1 \,|\, \boldsymbol{X} = \boldsymbol{x}\right)\Big)^{y} \\ &= q_0^{1-y}(\boldsymbol{x}) + q_1^{y}(\boldsymbol{x})\,, \end{aligned} \tag{128}$$

όπου $q_0(\boldsymbol{x})$ και $q_1(\boldsymbol{x})$ η πιθανότητα η τυχαία μεταβλητή Y να πάρει την τιμή 0 και την τιμή 1, αντίστοιχα, υπό την δέσμευση ότι η τυχαία μεταβλητή $\boldsymbol{X}$ παίρνει την τιμή $\boldsymbol{x}$. Αυτού του είδους τα προβλήματα ονομάζονται προβλήματα ταξινόμησης (classification problems) γιατί θεωρούμε ότι η τιμή της μεταβλητής Y δίνει την πληροφορία η μεταβλητή $\boldsymbol{x}$ να ανήκει στην κλάση '0' ή στην κλάση '1'.

Θεωρούμε ότι οι q_0 και q_1 είναι άγνωστες και τις μοντελοποιούμε με τις συναρτήσεις

$$p_0(\boldsymbol{x};\boldsymbol{\beta}) = \frac{1}{1+e^{\boldsymbol{x}^\top\boldsymbol{\beta}}}\,, \tag{129}$$

$$p_1(\boldsymbol{x};\boldsymbol{\beta}) = \frac{e^{\boldsymbol{x}^\top\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}^\top\boldsymbol{\beta}}}\,. \tag{130}$$

Αυτού του είδους η παλινδρόμηση ονομάζεται λογιστική παλινδρόμηση γιατί η συνάρτηση p_0 ονομάζεται λογιστική συνάρτηση. Παρατηρήστε ότι,

$$\log\frac{p_1(\boldsymbol{x})}{1-p_1(\boldsymbol{x})} = \boldsymbol{x}^\top\boldsymbol{\beta}\,. \tag{131}$$

Ο λόγος των πιθανοτήτων ονομάζεται λόγος απόδοσης ή log-odds, και δίνει την απόδοση όταν κάποιος στοιχηματίζει σε ένα ενδεχόμενο με πιθανότητα p_1 να συμβεί και $1-p_1$ να μη συμβεί.

Σε αντίθεση με την γραμμική παλινδρόμηση, όπου ελαχιστοποιούμε το άθροισμα των τετραγώνων των σφαλμάτων, στην λογιστική παλινδρόμηση μεγιστοποιούμε την ποσότητα,

$$\begin{aligned} L(\boldsymbol{\beta}\,|\,\boldsymbol{x}) &\coloneqq \mathbb{P}\left(Y_1 = y_1,\ldots,Y_n = y_n\,|\,\boldsymbol{X}_1 = \boldsymbol{x}_1,\ldots,\boldsymbol{X} = \boldsymbol{x}_n;\boldsymbol{\beta}\right) \\ &= \prod_{i=1}^{n}\mathbb{P}\left(Y = y_i\,|\,\boldsymbol{X} = \boldsymbol{x}_i;\boldsymbol{\beta}\right) \\ &= \prod_{i=1}^{n} p_0^{1-y}(\boldsymbol{x};\boldsymbol{\beta})\,p_1^{y}(\boldsymbol{x};\boldsymbol{\beta})\,. \end{aligned} \tag{132}$$

Η συνάρτηση L ονομάζεται συνάρτηση πιθανοφάνειας (likelihood function). Μεγιστοποιώντας την ως προς $\boldsymbol{\beta}$ μεγιστοποιούμε την πιθανοφάνεια των δεδομένων. Συνήθως δουλεύουμε με τον λογάριθμο της πιθανοφάνειας

$$\begin{aligned} \ell(\boldsymbol{\beta}\,|\,\boldsymbol{x}) &= \log L(\boldsymbol{\beta}\,|\,\boldsymbol{x}) \\ &= \sum_{i=1}^{n} y_i\boldsymbol{x}^\top\boldsymbol{\beta} - \log\left(1+e^{\boldsymbol{x}^\top\boldsymbol{\beta}}\right)\,. \end{aligned} \tag{133}$$

Εφόσον η log είναι αύξουσα συνάρτηση, η συνάρτησης πιθανοφάνειας παίρνει μέγιστο στο ίδιο σημείο με τον λογάριθμο της πιθανοφάνειας. Ορίζουμε την εκτιμήτρια του $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} \coloneqq \arg\max_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta}\,|\,\boldsymbol{x})\,. \tag{134}$$

Η συνάρτηση ℓ δεν είναι γραμμική ως προς τις παραμέτρους $\boldsymbol{\beta}$ και δεν μπορούμε να βρούμε κλειστό τύπο για την τιμή της εκτιμήτριας. Μπορούμε να βρούμε το μέγιστο της ℓ χρησιμοποιώντας κατάλληλους αλγορίθμους βελτιστοποίησης.

Έχοντας βρει τιμή για την $\hat{\boldsymbol{\beta}}$, για ένα καινούριο $\boldsymbol{x}_{new}$ που δεν ανήκει στο αρχικό σύνολο δεδομένων, μπορούμε να βρούμε την πιθανότητα η Y να ανήκει στην κλάση 0 ή 1. Εάν θέλουμε να απαντήσουμε οριστικά σε ποια κλάση ανήκει το $\boldsymbol{x}_{new}$, χρησιμοποιούμε την τιμή της $\hat{f}$,

$$\hat{f}(\boldsymbol{x}) = \begin{cases} 0, & p_1(\boldsymbol{x}) \le \frac{1}{2} \\ 1, & p_1(\boldsymbol{x}) > \frac{1}{2} \end{cases}\,, \tag{135}$$

ή εναλλακτικά,

$$\hat{f}(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x}^\top\boldsymbol{\beta} \ge 0 \\ 1, & \boldsymbol{x}^\top\boldsymbol{\beta} \ge 1 \end{cases}\,. \tag{136}$$

Το επίπεδο $\boldsymbol{x}^\top\boldsymbol{\beta} = 0$ ονομάζεται *σύνορο απόφασης* (decision boundary).

9.3 Περισσότερες από δύο κλάσεις

Στην περίπτωση που τα δεδομένα y_i παίρνουν τιμές $\{0, 1, \ldots, k\}$, μοντελοποιούμε τις συναρτήσεις πιθανότητας

$$q_i(\boldsymbol{x}) = \mathbb{P}\left(Y = i \mid \boldsymbol{X} = \boldsymbol{x}\right), \quad i = 0, \ldots, k, \tag{137}$$

με τις συναρτήσεις,

$$p_0(\boldsymbol{x}; \boldsymbol{\beta}) = \frac{1}{1 + \sum_{j=1}^{k} e^{\boldsymbol{x}^\top \boldsymbol{\beta}_j}}, \tag{138}$$

$$p_i(\boldsymbol{x}; \boldsymbol{\beta}) = \frac{e^{\boldsymbol{x}^\top \boldsymbol{\beta}_j}}{1 + \sum_{j=1}^{k} e^{\boldsymbol{x}^\top \boldsymbol{\beta}_j}}, \quad i = 1, \ldots, k, . \tag{139}$$

k διανύσματα παραμέτρων $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$.

10 Εβδομάδα 10/13

10.1 Διαστήματα εμπιστοσύνης: το πιο απλό μοντέλο

Θεωρούμε το σύνολο δεδομένων $\{\boldsymbol{x}_i\}_{i=1}^n$. Υποθέτουμε ότι κάθε $\boldsymbol{x}_i$ είναι πραγματοποίηση μιας τυχαίας μεταβλητής $\boldsymbol{X}_i$ όπου κάθε $\boldsymbol{X}_i$ ακολουθεί την ίδια κατανομή $p_{\boldsymbol{X}}$ με μέση τιμή $\boldsymbol{\mu}$ και διακύμανση $\sigma^2 I_n$. Θέλουμε να μάθουμε την τιμή της μέσης τιμής,

$$\boldsymbol{\beta} = \mathbb{E}_{p_{\boldsymbol{X}}}\left[\boldsymbol{X}\right]. \tag{140}$$

Θα εκτιμήσουμε την $\boldsymbol{\beta}$ με την εκτιμήτρια,

$$\hat{\boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i. \tag{141}$$

Για ένα συγκεκριμένο σύνολο δεδομένων η $\hat{\boldsymbol{\beta}}$ είναι ένα διάνυσμα. Αν πάρουμε ένα άλλο σύνολο δεδομένων η τιμή θα αλλάξει. Γενικά, μπορούμε να θεωρήσουμε το σύνολο $\{\boldsymbol{X}_i\}_{i=1}^n$ και να το θεωρήσουμε ως τυχαίο. Σε αυτή την περίπτωση η $\hat{\boldsymbol{\beta}}$ είναι η τυχαία μεταβλητή,

$$\hat{\boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i. \tag{142}$$

Μας ενδιαφέρει να βρούμε την μέση τιμή και την διασπορά της εκτιμήτριας.

Θεώρημα 10.1 (μέση τιμή και διασπορά της εκτιμήτριας)**.** *Η μέση τιμή της εκτιμήτριας (141),*

$$\mathbb{E}_{p_{\boldsymbol{X}}}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}, \tag{143}$$

και η συνδιακύμανση της,

$$\mathrm{Cov}_{p_{\boldsymbol{X}}}\left[\hat{\boldsymbol{\beta}}\right] = \frac{\sigma^2}{N} I. \tag{144}$$

Άσκηση 10.1. *Αποδείξτε το θεώρημα 10.1.*

Για δεδομένο σύνολο δεδομένων $\{\boldsymbol{x}_i\}_{i=1}^n$ θέλουμε να βρούμε ένα διάστημα στο οποίο η πραγματική τιμή της παραμέτρους ανήκει με συγκεκριμένη πιθανότητα.

Θεώρημα 10.2 (προσεγγιστικό διαστήμα εμπιστοσύνης στον $\mathbb{R}$)**.** *Έστω ένα σύνολο τυχαίων μεταβλητών $\{X_i\}_{i=1}^n$ με $X_i \in \mathbb{R}$ και $S_X^2 = \frac{1}{n}\sum_{i=1}^{n-1}\left(X_i - \hat{\beta}\right)^2$ η εκτιμήτρια της διακύμανσης. Για $a \in (0, 1)$ και μεγάλες τιμές του n,*

$$\mathbb{P}\left(\left(\hat{\beta} - c\sqrt{\frac{S_X^2}{n}}, \hat{\beta} + c\sqrt{\frac{S_X^2}{n}}\right) \ni \beta\right) \approx a, \tag{145}$$

για $c = \Phi^{-1}(\frac{a+1}{2})$.

Απόδειξη. Από το κεντρικό οριακό θεώρημα ισχύει προσεγγιστικά ότι για μεγάλες τιμές του n

$$\hat{\beta} \sim \mathcal{N}(\beta, \frac{\sigma^2}{n})\,. \tag{146}$$

Επομένως,

$$\mathbb{P}\left(-c \leq \frac{\hat{\beta}-\beta}{\frac{\sigma}{\sqrt{n}}} \leq c\right) = \Phi(c) - \Phi(-c)\,. \tag{147}$$

Εάν θέλουμε $\Phi(c) - \Phi(-c) \approx a$ τότε $c = \Phi^{-1}(\frac{a+1}{2})$ και έχουμε

$$\mathbb{P}\left(\left(\hat{\beta} - c\frac{\sigma}{\sqrt{n}}, \hat{\beta} + c\frac{\sigma}{\sqrt{n}}\right) \ni \hat{\beta}\right) \approx a\,. \tag{148}$$

Αντικαθιστώντας το σ με την εκτιμήτρια S_X παίρνουμε το ζητούμενο. □

Θεώρημα 10.3 (προσεγγιστικό διαστήμα εμπιστοσύνης στον $\mathbb{R}^k$)**.** *Έστω ένα σύνολο τυχαίων μεταβλητών* $\{\boldsymbol{X}_i\}_{i=1}^n$ *με* $X_i \in \mathbb{R}^k$ *και* $S_{\boldsymbol{X}}^2 = \frac{1}{n}\sum_{i=1}^{n-1}\left(\boldsymbol{X}_i - \hat{\boldsymbol{\beta}}\right)^2$ *η εκτιμήτρια του πίνακα συνδιακύμανσης. Για* $a \in (0,1)$ *και μεγάλες τιμές του* n*,*

$$\mathbb{P}\left(\left(\hat{\beta}_j - c\sqrt{\frac{[S_{\boldsymbol{X}}^2]_{jj}}{n}}, \hat{\beta}_j + c\sqrt{\frac{[S_{\boldsymbol{X}}^2]_{jj}}{n}}\right) \ni \beta_j\right) \approx a\,, \tag{149}$$

για $c = \Phi^{-1}(\frac{a+1}{2})$.

Ορισμός 10.1 (Student κατανομή)**.** *Η συνάρτηση πυκνότητας πιθανότητας της Student κατανομής με* n *βαθμούς ελευθερίας ορίζεται ως,*

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n+1}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}\,. \tag{150}$$

Θεώρημα 10.4. *Ισχύει ότι*

$$\frac{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\frac{S_{\boldsymbol{X}}}{\sqrt{n}}} \sim t_{n-1}\,, \tag{151}$$

όπου t_n *η κατανομή Student με* n *βαθμούς ελευθερίας.*

Θεώρημα 10.5 (ακριβές διαστήμα εμπιστοσύνης στον $\mathbb{R}$)**.** *Έστω ένα σύνολο τυχαίων μεταβλητών* $\{X_i\}_{i=1}^n$ *με* $X_i \in \mathbb{R}$ *και* $S_X^2 = \frac{1}{n}\sum_{i=1}^{n-1}\left(X_i - \hat{\beta}\right)^2$ *η εκτιμήτρια της διακύμανσης. Για* $a \in (0,1)$ *και για κάθε* $n \in \mathbb{N}$ *με* $n > 2$*,*

$$\mathbb{P}\left(\left(\hat{\beta} - c\sqrt{\frac{S_X^2}{n}}, \hat{\beta} + c\sqrt{\frac{S_X^2}{n}}\right) \ni \beta\right) = a\,, \tag{152}$$

για $c = F^{-1}(\frac{1-a}{2})$ *και* F^{-1} *την συνάρτηση ποσοστημόριου της Student κατανομής με* $n-1$ *βαθμούς ελευθερίας.*

Το Θεώρημα 10.5 γενικεύεται στον $\mathbb{R}^k$ με τον ίδιο τρόπο που το θεώρημα 10.3 γενικεύει το θεώρημα 10.2.

10.2 Διαστήματα εμπιστοσύνης: γραμμική παλινδρόμηση

Στην πολλαπλή γραμμική παλινδρόμηση που μελετήσαμε στην παράγραφο 7.2, εμφανίζονται οι τυχαίες μεταβλητές $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{Y}} = A\hat{\boldsymbol{\beta}}$, και $\hat{Y}_{new} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}_{new}$.

Θεώρημα 10.6. *Για την εκτιμήτρια στην σχέση (114), ισχύει ότι*

$$\mathbb{E}\left[\hat{\boldsymbol{\beta}} \,|\, A\right] = \boldsymbol{\beta}\,, \tag{153}$$

και

$$\mathrm{Cov}\left[\hat{\boldsymbol{\beta}} \,|\, A\right] = \sigma^2 (A^\top A)^{-1}\,. \tag{154}$$

Απόδειξη. Για την δεσμευμένη μέση τιμή της εκτιμήτριας ισχύει ότι,

$$\begin{aligned} \mathbb{E}\left[\hat{\boldsymbol{\beta}} \,|\, A\right] &= \mathbb{E}\left[(A^\top A)^{-1} A^\top \boldsymbol{y} \,|\, A\right] \\ &= (A^\top A)^{-1} A^\top \mathbb{E}\left[\boldsymbol{y} \,|\, A\right] \\ &= (A^\top A)^{-1} A^\top A \boldsymbol{\beta} \\ &= \boldsymbol{\beta}\,. \end{aligned} \tag{155}$$

Για την πίνακα διασποράς της εκτιμήτριας ισχύει ότι

$$\begin{aligned} \mathrm{Cov}\left[\hat{\boldsymbol{\beta}} \,|\, A\right] &= \mathrm{Cov}\left[(A^\top A)^{-1} A^\top \boldsymbol{y} \,|\, A\right] \\ &= (A^\top A)^{-1} A^\top \,\mathrm{Cov}\left[\boldsymbol{y} \,|\, A\right] A (A^\top A)^{-1} \\ &= (A^\top A)^{-1} A^\top \,\sigma^2 I_n \, A (A^\top A)^{-1} \\ &= \sigma^2 \,(A^\top A)^{-1} A^\top A (A^\top A)^{-1} \\ &= \sigma^2 \,(A^\top A)^{-1}\,. \end{aligned} \tag{156}$$

□

Ερώτηση 10.1. *Αποδείξτε ότι* $\mathrm{Cov}\left[A\boldsymbol{y}\right] = A\,\mathrm{Cov}\left[\boldsymbol{y}\right] A^\top$.

Θεώρημα 10.7. *Για την τυχαία μεταβλητή* $\hat{\boldsymbol{Y}} = A\hat{\boldsymbol{\beta}}$ *ισχύει ότι,*

$$\mathbb{E}\left[\hat{\boldsymbol{Y}} \,|\, A\right] = A\boldsymbol{\beta}\,, \tag{157}$$

και

$$\mathrm{Cov}\left[\hat{\boldsymbol{Y}} \,|\, A\right] = \sigma^2 A (A^\top A)^{-1} A^\top\,. \tag{158}$$

Άσκηση 10.2. *Αποδείξτε το θεώρημα 10.7.*

Θεώρημα 10.8. *Για την τυχαία μεταβλητή* $\hat{Y}_{new} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}_{new}$ *ισχύει ότι,*

$$\mathbb{E}\left[\hat{Y}_{new} \,|\, A, \boldsymbol{X}_{new} = \boldsymbol{x}_{new}\right] = \boldsymbol{x}_{new}^\top \boldsymbol{\beta}\,, \tag{159}$$

και

$$\mathbb{V}\left[\hat{Y}_{new} \,|\, A, \boldsymbol{X}_{new} = \boldsymbol{x}_{new}\right] = \sigma^2 \boldsymbol{x}_{new}^\top (A^\top A)^{-1} \boldsymbol{x}_{new}\,. \tag{160}$$

Απόδειξη. Παρατηρήστε ότι το θεώρημα είναι όμοιο με το θεώρημα 10.7 με τη διαφορά ότι η μεταβλητή $\hat{Y}_{new}$ μία από τις συντεταγμένες του διανύσματος $\hat{\boldsymbol{Y}}$. □

Ορισμός 10.2 (standard error)**.** *Η εκτίμηση της τυπική απόκλιση μιας εκτιμήτριας* $\hat{\beta}$ *ονομάζεται* τυπικό σφάλμα *και συμβολίζεται με* $SE(\hat{\beta})$ *ή* $\sigma_{\hat{\beta}}$.

Το τυπικό σφάλμα της j συνιστώσας της εκτιμήτριας (114) δίνεται από τη σχέση (156) και την εκτιμήτρια για το $\widehat{\sigma^2}$ από το θεώρημα 7.2,

$$SE(\hat{\beta}_j) = \widehat{\sigma^2}\left[(A^\top A)^{-1}\right]_{jj}\,. \tag{161}$$

Το τυπικό σφάλμα της $\hat{Y}_{new}$ δίνεται από το θεώρημα 10.8,

$$SE(\hat{Y}_{new}) = \widehat{\sigma^2} \boldsymbol{x}_{new}^\top (A^\top A)^{-1} \boldsymbol{x}_{new}\,. \tag{162}$$

Θεώρημα 10.9 (διαστήματα εμπιστοσύνης)**.** *Ένα διάστημα εμπιστοσύνης για την πραγματική τιμή* β *δίνεται από την σχέση*

$$\mathbb{P}\left(\left(\hat{\beta}_j - SE(\hat{\beta}_j) F^{-1}\left(\frac{1-a}{2}\right), \hat{\beta}_j + SE(\hat{\beta}_j) F^{-1}\left(\frac{1-a}{2}\right)\right) \ni \beta_j\right) = a\,, \tag{163}$$

όπου F^{-1} *η συνάρτηση ποσοστημόριου της Student κατανομής με* $n-m-1$ *βαθμούς ελευθερίας.*

Ένα διάστημα εμπιστοσύνης για την πραγματική μέση τιμή της πρόβλεψης $\mathbb{E}\left[Y_{new} \,|\, \boldsymbol{x}_{new}\right] = \boldsymbol{\beta}^\top \boldsymbol{x}_{new}$ *δίνεται από τη σχέση,*

$$\mathbb{P}\left(\left(\hat{Y}_{new} - SE(\hat{Y}_{new})F^{-1}\left(\frac{1-a}{2}\right), \hat{Y}_{new} + SE(\hat{Y}_{new})F^{-1}\left(\frac{1-a}{2}\right)\right) \ni \boldsymbol{\beta}^\top \boldsymbol{x}_{new}\right) = a\,. \tag{164}$$

11 Εβδομάδα 11/13

11.1 Fisher Information

Ορισμός 11.1 (συνάρτηση σκορ)**.** *Έστω $p(x;\boldsymbol{\beta})$ μια συνάρτηση πυκνότητας πιθανότητας ή μια συνάρτηση μάζας πιθανότητας παραμετρικοποιημένη από το διάνυσμα $\boldsymbol{\beta}$. Η συνάρτηση σκορ (score function) ορίζεται ως,*

$$s(\boldsymbol{\beta};x) = \frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta})\,. \tag{165}$$

Πρόταση 11.1. *Η μέση τιμή της συνάρτηση σκορ είναι μηδέν.*

Απόδειξη.

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta})\right] &= \int \frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta}) p(x;\boldsymbol{\beta})\,\mathrm{d}x \\
&= \int \frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta}) p(x;\boldsymbol{\beta})\,\mathrm{d}x \\
&= \int \frac{1}{p(x;\boldsymbol{\beta})} \frac{\partial}{\partial \boldsymbol{\beta}} p(x;\boldsymbol{\beta})\,\mathrm{d}x \\
&= \int \frac{\partial}{\partial \boldsymbol{\beta}} p(x;\boldsymbol{\beta})\,\mathrm{d}x \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \int p(x;\boldsymbol{\beta})\,\mathrm{d}x \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} 1 = 0
\end{aligned} \tag{166}$$

□

Ορισμός 11.2 (Fisher Information)**.** *Ο πίνακας Fisher Information της $p(x;\boldsymbol{\beta})$ ορίζεται ως,*

$$I(\beta) = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta}) \left(\frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta})\right)^\top\right]\,. \tag{167}$$

Παρατήρηση 11.1. *Παρατηρήστε ότι*

$$I(\beta) = \mathbb{V}\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log p(x;\boldsymbol{\beta})\right]\,. \tag{168}$$

Πρόταση 11.2.

$$I(\beta) = -\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\beta}^2} \log p(x;\boldsymbol{\beta})\right]\,. \tag{169}$$

όπου $\dfrac{\partial^2}{\partial \boldsymbol{\beta}^2}$ *ο Εσσιανός πίνακας με το i,j στοιχείο ίσο με* $\frac{\partial^2}{\partial \beta_i \partial \beta_j}$.

Απόδειξη.

$$\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\beta}^2}\log p(x;\boldsymbol{\beta})\right] &= -\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}}\frac{\frac{\partial}{\partial \boldsymbol{\beta}}\log p(x;\boldsymbol{\beta})}{p(x;\boldsymbol{\beta})}\right] \\
&= -\mathbb{E}\left[\frac{p(x;\boldsymbol{\beta})\frac{\partial^2}{\partial \boldsymbol{\beta}^2}p(x;\boldsymbol{\beta}) - \left(\frac{\partial}{\partial \boldsymbol{\beta}}p(x;\boldsymbol{\beta})\right)^2}{p^2(x;\boldsymbol{\beta})}\right] \\
&= -\mathbb{E}\left[\frac{\frac{\partial^2}{\partial \boldsymbol{\beta}^2}p(x;\boldsymbol{\beta})}{p(x;\boldsymbol{\beta})}\right] + \mathbb{E}\left[\left(\frac{\frac{\partial}{\partial \boldsymbol{\beta}}p(x;\boldsymbol{\beta})}{p(x;\boldsymbol{\beta})}\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{\frac{\partial}{\partial \boldsymbol{\beta}}p(x;\boldsymbol{\beta})}{p(x;\boldsymbol{\beta})}\right)^2\right] \\
&= I(\boldsymbol{\beta})\,,
\end{aligned} \tag{170}$$

διότι

$$\begin{aligned}
\mathbb{E}\left[\frac{\frac{\partial^2}{\partial \boldsymbol{\beta}^2}p(x;\boldsymbol{\beta})}{p(x;\boldsymbol{\beta})}\right] &= \int \frac{\frac{\partial^2}{\partial \boldsymbol{\beta}^2}p(x;\boldsymbol{\beta})}{p(x;\boldsymbol{\beta})}p(x;\boldsymbol{\beta})\,\mathrm{d}x \\
&= \frac{\partial^2}{\partial \boldsymbol{\beta}^2}\int p(x;\boldsymbol{\beta})\,\mathrm{d}x = 0\,.
\end{aligned} \tag{171}$$

□

Παράδειγμα 11.1 (Fisher Information για την Bernoulli κατανομή). *Η συνάρτηση πυκνότητας πιθανότητας της Bernoulli κατανομής δίνεται ως*

$$p(x,\beta) = \beta^x(1-\beta)^{1-x}\,, \tag{172}$$

και ο λογάριθμος της δίνεται ως,

$$\ell(\beta;x) = x\log\beta + (1-x)\log(1-\beta)\,. \tag{173}$$

Η δεύτερη παράγωγος της ℓ είναι

$$\ell''(\beta;x) = -\frac{x}{\beta^2} - \frac{1-x}{1-\beta^2}. \tag{174}$$

Ο Fisher Information δίνεται ως,

$$\begin{aligned}
I(\beta) &= -\mathbb{E}\left[-\frac{X}{\beta^2} + \frac{1-X}{1-\beta^2}\right] \\
&= \frac{1}{\beta^2}\mathbb{E}\left[X\right] + \frac{1-\mathbb{E}\left[X\right]}{1-\beta^2} \\
&= \frac{\beta}{\beta^2} + \frac{1-\beta}{(1-\beta)^2} \\
&= \frac{1}{\beta(1-\beta)}\,,
\end{aligned} \tag{175}$$

διότι $\mathbb{E}\left[X\right] = \beta$.

Παράδειγμα 11.2 (Fisher Information για την κανονική κατανομή). *Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής δίνεται ως*

$$p(x;\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\tau}}\mathrm{e}^{-\frac{1}{2\tau}(x-\mu)^2}\,, \tag{176}$$

όπου $\boldsymbol{\beta} = (\mu, \tau)$ και upptau $= \sigma^2$. Η συνάρτηση score

$$s(\boldsymbol{\beta}; x) = -\frac{1}{2}\log 2\pi + \frac{1}{2}\log\tau - \frac{1}{2\tau}(x-\mu)^2 \,. \tag{177}$$

Η δεύτερη παράγωγος της s ως προς μ δίνεται από

$$\frac{\partial^2 s}{\partial \mu^2}(\boldsymbol{\beta}; x) = -\frac{1}{\tau}\,, \tag{178}$$

η δεύτερη παράγωγος ως προς τ δίνεται από

$$\frac{\partial^2 s}{\partial \mu^2}(\boldsymbol{\beta}; x) = \frac{1}{2\tau^2} - \frac{(x-\mu)^2}{\tau^3}\,, \tag{179}$$

και η μικτή παράγωγος δίνεται από,

$$\frac{\partial^2 s}{\partial \mu \partial \tau} = (\boldsymbol{\beta}; x) = \frac{1}{2\tau^2} - \frac{(x-\mu)^2}{\tau^3}\,, \tag{180}$$

Εφόσον $\mathbb{E}\left[X-\mu\right] = 0$ και $\mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$, ο πίνακας Fisher Information δίνεται ως,

$$I(\boldsymbol{\beta}) = \begin{bmatrix} \frac{1}{\tau} & 0 \\ 0 & \frac{1}{2\tau^2} \end{bmatrix}. \tag{181}$$

Πρόταση 11.3. *Έστω n ανεξάρτητες τυχαίες μεταβλητές $\{X_i\}_{i=1}^n$ με $X_i \sim p(\cdot; \boldsymbol{\beta})$. Εάν I ο Fisher Information της p και I_n ο Fisher Information της $p_n(x_1, x_2, \ldots, x_n; \boldsymbol{\beta})$ τότε*

$$I_n(\boldsymbol{\beta}) = nI(\boldsymbol{\beta})\,. \tag{182}$$

Απόδειξη. Εφόσον οι μεταβλητές είναι ανεξάρτητες, ισχύει ότι

$$p_n(x_1, x_2, \ldots, x_n; \boldsymbol{\beta}) = \prod_{i=1}^{n} p(x_i)\,, \tag{183}$$

και

$$\log p_n(x_1, x_2, \ldots, x_n; \boldsymbol{\beta}) = \sum_{i=1}^{n} \log p(x_i)\,. \tag{184}$$

Επομένως,

$$\begin{aligned} I_n(\boldsymbol{\beta}) &= -\mathbb{E}\left[\tfrac{\partial^2}{\partial \boldsymbol{\beta}^2}\log p_n(x_1, x_2, \ldots, x_n; \boldsymbol{\beta})\right] \\ &= \sum_{i=1}^{n} -\mathbb{E}\left[\tfrac{\partial^2}{\partial \boldsymbol{\beta}^2}\log\log p(x_i)\right] \\ &= \sum_{i=1}^{n} I(\boldsymbol{\beta}) \\ &= nI(\boldsymbol{\beta})\,. \end{aligned} \tag{185}$$

□

Θεώρημα 11.1 (Cramér-Rao). *Για $\hat{\beta}$ αμερόληπτη εκτιμήτρια της β χρησιμοποιώντας n δείγματα, ισχύει ότι,*

$$\mathbb{V}\left[\hat{\beta}\right] \geq \frac{1}{I_n(\beta)}. \tag{186}$$

Θεώρημα 11.2 (ασυμπτωτική κατανομή). *Η ασυμπτωτική κατανομή της αμερόληπτης εκτιμήτριας $\hat{\beta}$ της β χρησιμοποιώντας n δείγματα είναι η κανονική κατανομή με μέση τιμή β και διακύμανση $I^{-1}(\beta)$.*

Παρατήρηση 11.2 (διάστημα εμπιστοσύνης). *Από το θεώρημα 11.2, για μεγάλο n μπορούμε να κατασκευάσουμε ένα προσεγγιστικό διάστημα εμπιστοσύνης,*

$$\left(\hat{\beta} - cI_n^{-1}(\beta), \hat{\beta} + cI_n^{-1}(\beta)\right) \ni \beta\,. \tag{187}$$

11.2 Διαστήματα εμπιστοσύνης: λογιστική παλινδρόμηση

Στην παράγραφο αυτή θα κατασκευάσουμε θα υπολογίσουμε τον Fisher Information και σύμφωνα με την παρατήρηση 11.2 μπορούμε να κατασκευάσουμε ένα προσεγγιστικό διάστημα εμπιστοσύνης για την παράμετρο β του λογιστικού μοντέλου. Από την σχέση (132), ο λογάριθμος της πιθανοφάνειας για την λογιστική παλινδρόμηση δίνεται από την σχέση,

$$\begin{aligned}\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n} y_i \log \frac{p_1(\boldsymbol{x}_i;\boldsymbol{\beta})}{1-p_1(\boldsymbol{x}_i;\boldsymbol{\beta})} + \log(1-p_1(\boldsymbol{x}_i;\boldsymbol{\beta})) \\ &= \sum_{i=1}^{n} y_i \log \mathrm{e}^{\boldsymbol{x}_i^\top \boldsymbol{\beta}} + \log \frac{1}{1+\mathrm{e}^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}} \\ &= \sum_{i=1}^{n} y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \log\left(1+\mathrm{e}^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\right) .\end{aligned} \tag{188}$$

Η πρώτη παράγωγος ως προς $\boldsymbol{\beta}$ υπολογίζεται ως,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\beta}}\,\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n} y_i \boldsymbol{x}_i - \left(1+\mathrm{e}^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\right)^{-1} \mathrm{e}^{\boldsymbol{x}_i^\top \boldsymbol{\beta}} \boldsymbol{x}_i \\ &= \sum_{i=1}^{n} \big(y_i - p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)\, \boldsymbol{x}_i .\end{aligned} \tag{189}$$

Υπολογίζουμε,

$$\begin{aligned}I(\beta) &= \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}}\,\ell(\boldsymbol{\beta}) \left(\frac{\partial}{\partial \boldsymbol{\beta}}\,\ell(\boldsymbol{\beta})\right)^\top\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{n} \big(y_i - p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)\, \boldsymbol{x}_i\right)\left(\sum_{i=1}^{n} \big(y_i - p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)\, \boldsymbol{x}_i^\top\right)\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \big(y_i - p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)\big(y_j - p_1(\boldsymbol{x}_j;\boldsymbol{\beta})\big)\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \mathbb{E}\left[\big(y_i - p_1(\boldsymbol{x}_j;\boldsymbol{\beta})\big)\big(y_j - p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)\right] .\end{aligned} \tag{190}$$

Κάθε μεταβλητή Y_i ακολουθεί Bernoulli κατανομή με παράμετρο $p_1(\boldsymbol{x}_i;\boldsymbol{\beta})$. Επομένως, ισχύει ότι $\mathbb{E}\left[Y_i\right] = p_1(\boldsymbol{x}_i;\boldsymbol{\beta})$ και η πάνω σχέση γράφεται ως,

$$I(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \operatorname{Cov}\left[(Y_1,\ldots,Y_n) \,|\, \boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\right] . \tag{191}$$

Οι τυχαίες μεταβλητές είναι ανεξάρτητες, επομένως ο πίνακας συνδιακύμανσης είναι διαγώνιος, και ο πίνακας Fisher Information γράφεται ως,

$$\begin{aligned}I(\beta) &= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \mathbb{V}\left[Y_i \,|\, \boldsymbol{x}_i\right] \\ &= \sum_{i=1}^{n} p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big(1-p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)\, \boldsymbol{x}_i \boldsymbol{x}_i^\top ,\end{aligned} \tag{192}$$

διότι η διακύμανση της Y είναι ίση με $\mathbb{V}\left[Y_i\right] = p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big(1-p_1(\boldsymbol{x}_i;\boldsymbol{\beta})\big)$.

11.3 Γραμμική παλινδρόμηση: εκτιμήτρια μέγιστης πιθανοφάνειας

Στην παράγραφο αυτή θα κατασκευάσουμε θα υπολογίσουμε τον Fisher Information και σύμφωνα με την παρατήρηση 11.2 μπορούμε να κατασκευάσουμε ένα προσεγγιστικό διάστημα εμπιστοσύνης για την παράμετρο β του γραμμικού μοντέλου που μελετήσαμε στην παράγραφο 7.1. Για να υπολογίσουμε την συνάρτηση πιθανοφάνειας χρειαζόμαστε μία επιπρόσθετη υπόθεση για το μοντέλο. Υποθέτουμε ότι οι παρατηρήσεις Y ακολουθούν κατανομή

$$Y \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{x}^\top \boldsymbol{\beta}, \sigma^2\right) . \tag{193}$$

Με την υπόθεση αυτή, η υπόθεση της γραμμικής παλινδρόμησης 106 εξακολουθεί να ισχύει. Η συνάρτηση πιθανοφάνειας γράφεται ως,

$$L\left(\boldsymbol{\beta}, \tau; \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n\right) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{1}{2\tau}(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2\right) , \tag{194}$$

όπου $\tau = \sigma^2$. Ο λογάριθμος της συνάρτησης πιθανοφάνειας γράφεται ως

$$\begin{aligned} \ell\left(\boldsymbol{\beta}, \tau; \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n\right) &= c + n \log \tau + \tau^{-1} \sum_{i=1}^n \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 \\ &= c + n \log \tau + \tau^{-1} \mathcal{E}(\beta) . \end{aligned} \tag{195}$$

όπου c σταθερός όρος που δεν εξαρτάται από τις $\boldsymbol{\beta}$ και τ. Παραγωγίζοντας την ℓ ως προς $\boldsymbol{\beta}$ και θέτοντας ίση με τον μηδέν, καταλήγουμε στην ίδια εκτιμήτρια με την εκτιμήτρια ελαχίστων τετραγώνων (114). Παραγωγίζοντας ως προς τ η εκτιμήτρια του τ δίνεται από την σχέση

$$\tau = \frac{1}{n} \mathcal{E}(\beta) \tag{196}$$

η οποία διαφέρει από την εκτιμήτρια (115) ως προς την κανονικοποίηση. Ισχύει ότι η εκτιμήτρια ελαχίστων τετραγώνων είναι αμερόληπτη ενώ η εκτιμήτρια μέγιστης πιθανοφάνειας είναι ασυμπτωτικά αμερόληπτη.

Άσκηση 11.1. *Δείξτε ότι ο πίνακας Fisher Information για τις παραμέτρους της γραμμικής παλινδρόμησης δίνεται από την σχέση,*

$$I(\beta, \tau) = \left[\begin{array}{c|c} \tau^{-1} X^\top X & 0_{m+1,m+1} \\ \hline 0_{m+1,m+1} & \frac{1}{2}\tau^{-2} \end{array}\right] . \tag{197}$$

12 Εβδομάδα 12/13

12.1 Χρονοσειρές

Ορισμός 12.1 (Χρονοσειρά - Τυχαία διαδικασία)**.** *Ως χρονοσειρά ορίζουμε μια ακολουθία τυχαίων μεταβλητών $X_t(\omega)$, $t = 0, 1, \ldots$, όπου ο δείκτης t έχει την έννοια του χρόνου.*

Οι τυχαίες μεταβλητές, για κάθε χρονική στιγμή, ακολουθούν μια κατανομή, Έστω p_t. Γενικά, οι X_t δεν είναι ανεξάρτητες μεταξύ τους. Επίσης, η απο κοινού κατανομή για οποιαδήποτε συλλογή k μεταβλητών $(X_{t_1}, \ldots, X_{t_k})$ διαφέρει από την κατανομή μιας διαφορετικής συλλογής k μεταβλητών $(X_{s_1}, \ldots, X_{s_k})$.

Ο παρακάτω ορισμός, δίνει μια υποκατηγορία χρονοσειρών που θα μας απασχολήσουν εδώ.

Ορισμός 12.2 (στάσιμες χρονοσειρές)**.** *Μια χρονοσειρά ονομάζεται ασθενώς στάσιμη εάν η μέση τιμή της κάθε τυχαίας μεταβλητής X_t είναι σταθερή ως προς τον χρόνο,*

$$\mathbb{E}_{p_t}[X_t] = \int x \, p_t(x) \, \mathrm{d}x = \mu . \tag{198}$$

Μια χρονοσειρά ονομάζεται αυστηρώς στάσιμη εάν η κατανομή του διανύσματος $(X_{t_1}, \ldots, X_{t_k})$ είναι ίδια με την κατανομή του διανύσματος $(X_{t_1+\tau}, \ldots, X_{t_k+\tau})$ για κάθε $k \in \mathbb{N}$ και κάθε $\tau \in \mathbb{Z}$. Η παράμετρος τ λέγεται καθυστέρηση και παίρνει και αρνητικές τιμές.

Ορισμός 12.3 (αυτοδιασπορά - αυτοδιακύμανση)**.** *Η αυτοδιασπορά μιας χρονοσειράς X_t δίνεται από την σχέση*

$$\gamma_{s,t} = \mathrm{Cov}_{p_{s,t}}[X_s, X_t] = \int (x - \mu_s)(y - \mu_t) \, p_{s,t}(x, y) \, \mathrm{d}x \, \mathrm{d}y . \tag{199}$$

Η αυτοσυσχέτιση δίνεται από την σχέση

$$\rho_{s,t} = \frac{\gamma_{s,t}}{\sqrt{\gamma_{s,s}\gamma_{t,t}}} . \tag{200}$$

Για μία αυστηρά στάσιμη χρονοσειρά η αυτοδιασπορά και η αυτοσυσχέτιση εξαρτώνται μόνο από την διαφορά $\tau = t - s$,

$$\gamma_\tau = \mathrm{Cov}_{p_{t,t+\tau}}\left[X_t, X_{t+\tau}\right] , \quad \forall t \in \mathbb{N} , \tag{201}$$

και

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} . \tag{202}$$

12.2 Μοντέλα χρονοσειρών

Η ιδέα είναι να ορίσουμε κάποια μοντέλα χρονοσειρών έτσι ώστε δεδομένων κάποιον παρατηρήσεων $\{x_t\}_{t=0}^{n}$ να μπορούμε να μπορούμε να βρούμε τις τιμές των παραμέτρων του μοντέλου και να κάνουμε προβλέψεις για μελλοντικές χρονικές στιγμές.

12.2.1 white noise

Λευκό θόρυβο (white noise) ονομάζουμε οποιαδήποτε χρονοσειρά W_t με τις ιδιότητες

$$\mathbb{E}\left[W_t\right] = 0 , \qquad \mathbb{V}\left[W_t\right] = \sigma_2 , \quad \forall t \in \mathbb{N} , \tag{203}$$

και γράφουμε $W_t \sim WN(\sigma^2)$.

Παράδειγμα 12.1 (συνάρτηση αυτοσυσχέτισης)**.** *Εφόσον οι W_t είναι ανεξάρτητες,*

$$\gamma_\tau = \mathrm{Cov}\left[X_t, X_{t+\tau}\right] = \begin{cases} 0, & \tau \neq 0, \\ \sigma^2, & \tau = 0, \end{cases} \tag{204}$$

και

$$\rho_\tau = \begin{cases} 0\,, & \tau \neq 0\,, \\ 1\,, & \tau = 0\,. \end{cases} \tag{205}$$

12.2.2 moving average

Ορίζουμε το moving average μοντέλο τάξης p για $p > 0$, $MA(p)$, ως,

$$X_t = W_t + a_1 W_{t-1} + \cdots + a_p W_{t-p} = \sum_{i=1}^{p} a_i W_{t-i} + W_t \,, \tag{206}$$

όπου $W_t \sim WN(\sigma^2)$.

Παράδειγμα 12.2 (συνάρτηση αυτοσυσχέτισης)**.** *Για $\tau = 0$,*

$$\begin{aligned} \mathbb{V}\left[X_t\right] &= \mathbb{V}\left[\sum_{i=1}^{p} a_i W_{t-i} + W_t\right] \\ &= \sum_{i=1}^{p} a_i^2 \mathbb{V}\left[W_{t-i}\right] + \mathbb{V}\left[W_t\right] \\ &= \sigma^2 \left(1 + \sum_{i=1}^{p} a_i^2\right) \end{aligned} \tag{207}$$

Για την αυτοδιασπορά, ξεκινάμε τον υπολογισμό για το μοντέλο $p = 1$. Παρατηρήστε ότι για $\tau > 1$ η αυτοδιασπορά είναι μηδέν. Για $\tau = 1$,

$$\begin{aligned} \mathrm{Cov}\left[X_{t+1} X_t\right] &= \mathbb{E}\left[(a_1 W_t + W_{t+1})(a_1 W_{t-1} + W_t)\right] \\ &= a_1^2 \cancelto{0}{\mathbb{E}\left[W_t W_{t-1}\right]} + a_1 \mathbb{E}\left[W_t^2\right] + a_1 \cancelto{0}{\mathbb{E}\left[W_{t+1} W_{t-1}\right]} + \cancelto{0}{\mathbb{E}\left[W_{t+1} W_t\right]} \\ &= a_1^2 \sigma^2 \,, \end{aligned} \tag{208}$$

διότι οι W_t είναι ανεξάρτητες. Επομένως η συνάρτηση αυτοδιασποράς για $p = 1$ δίνεται από την σχέση,

$$\gamma_\tau = \begin{cases} \sigma^2\left(1 + a_1^2\right), & \tau = 0, \\ a_1 \sigma^2, & \tau = 1, \\ 0, & \tau > 1\,. \end{cases} \tag{209}$$

Για $p > 1$ και $\tau > 0$,

$$
\begin{aligned}
\operatorname{Cov}\left[X_{t+\tau}X_t\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{p} a_i W_{t+\tau-i} + W_{t+\tau}\right)\left(\sum_{i=1}^{p} a_i W_{t-i} + W_t\right)\right] \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j \mathbb{E}\left[W_{t+\tau-i}W_{t-j}\right] + \sum_{i=1}^{p} a_i \mathbb{E}\left[W_{t+\tau-i}W_t\right] + \sum_{i=1}^{p} a_i \mathbb{E}\left[W_{t+\tau}W_{t-i}\right] + \cancelto{0}{\mathbb{E}\left[W_{t+\tau}W_t\right]}
\end{aligned}
\tag{210}
$$

Για να συνεχίσουμε τον υπολογισμό παρατηρούμε ότι $\mathbb{E}\left[W_s W_t\right] \neq 0$ όταν $t = s$. Ο τρίτος όρος στην σχέση (210) είναι μηδέν διότι είναι μη μηδέν όταν $\tau = -i$, έχουμε όμως υποθέσει ότι $\tau > 0$.

Ο δεύτερος όρος είναι μη μηδενικός όταν $i = \tau$ και είναι ίσος με $a_\tau \sigma^2$.

Ο πρώτος όρος είναι μη μηδενικός όταν $j = i - \tau$. Επειδή $j > 0$ πρέπει $i > \tau$ και ο πρώτος όρος είναι ίσος με

$$
\sigma^2 \sum_{i=\tau+1}^{p} a_i a_{i-\tau}\,. \tag{211}
$$

Επομένως, η συνάρτηση αυτοσυσχέτισης είναι ίση με

$$
\gamma_\tau = \begin{cases} \sigma^2\left(1 + a_1^2\right), & \tau = 0, \\ \sigma^2\left(a_\tau + \sum_{i=\tau+1}^{p} a_i a_{i-\tau}\right), & 1 \leq \tau \leq p, \\ 0, & \tau > 1\,. \end{cases} \tag{212}
$$

12.2.3 autoregressive

Ορίζουμε το autoregressive μοντέλο τάξης p για $p > 0$, $AR(p)$, ως,

$$
X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + W_t = \sum_{i=1}^{p} a_i X_{t-i} + W_t\,, \tag{213}
$$

όπου $W_t \sim WN(\sigma^2)$. Για $p = 1$ και $a_1 = 1$,

$$
X_t = X_{t-1} + W_t\,, \tag{214}
$$

το μοντέλο ονομάζεται *τυχαίος περίπατος* (random walk).

Το μοντέλο ονομάζεται autoregressive (αυτο-παλινδρόμηση) διότι χρησιμοποιεί τιμές από το παρελθόν της ίδιας της χρονοσειράς, και όχι κάποια άλλη εξωτερική επεξηγηματική μεταβλητή, για να εξηγήσει την τιμή της χρονοσειράς την χρονική τιμή t.

Παράδειγμα 12.3 (συνάρτηση αυτοσυσχέτισης)**.** *Υπολογίζουμε αρχικά την μέση τιμή της X_t,*

$$
\mathbb{E}\left[X_t\right] = \sum_{i=1}^{p} a_i \mathbb{E}\left[X_{t-i}\right] + \mathbb{E}\left[W_t\right]\,. \tag{215}
$$

Υπό την υπόθεση ότι η X_t είναι ισχυρά στάσιμη με μέση τιμή μ και διασπορά σ_X^2, έχουμε ότι, $\mu = \mu \sum_{i=1}^{p} a_i$ και άρα $\mu = 0$.

Δουλεύοντας παρόμοια για την διασπορά του X_t,

$$
\begin{aligned}
\mathbb{V}\left[X_t\right] &= a^2 \mathbb{V}\left[X_{t-1}\right] + \mathbb{V}\left[W_t\right] \Rightarrow \\
\sigma_X^2 &= a^2 \sigma_X^2 + \sigma^2 \Rightarrow \\
\sigma_X^2 &= \frac{\sigma^2}{1 - a^2}\,.
\end{aligned}
\tag{216}
$$

για $a \in (0, 1)$.

Για την συνδιακύμανση έχουμε για $\tau > 0$,

$$
\begin{aligned}
\operatorname{Cov}\left[X_t, X_{t+\tau}\right] &= \mathbb{E}\left[X_t X_{t+\tau}\right] - \mu^2 \\
&= \mathbb{E}\left[X_t X_{t+\tau}\right] \\
&= \mathbb{E}\left[X_t(a X_{t+\tau-1+W_{t-1}})\right] \\
&= a\mathbb{E}\left[X_t X_{t+\tau-1}\right] + \mathbb{E}\left[X_t W_{t-1}\right] \\
&= a\mathbb{E}\left[X_t X_{t+\tau-1}\right]
\end{aligned}
\tag{217}
$$

όπου ο όρος $\mathbb{E}[X_t W_{t-1}]$ είναι μηδέν διότι η X_t είναι ανεξάρτητη από τη W_{t-1}. Δουλεύοντας όμοια με τον όρο $\mathbb{E}[X_t X_{t+\tau-1}]$ και κάνοντας την αναδρομή για τ βήματα, έχουμε

$$\mathrm{Cov}\left[X_t, X_{t+\tau}\right] = a^\tau \mathbb{E}\left[X_t^2\right] = a^\tau \sigma_X^2 = \frac{\sigma^2}{1-a^2}\,. \tag{218}$$

Τελος, η συνάρτηση αυτοσυσχέτισης δίνεται από την σχέση $\rho_\tau = a^\tau$ για $a \in (0,1)$.

12.2.4 autoregressive moving average

Το μοντέλο $ARMA(p,q)$ τάξης p και q είναι συνδυασμός των $AR(p)$ και $MA(q)$ μοντέλων,

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{p} b_i W_{t-i} + W_t\,, \tag{219}$$

12.3 Εκτιμήτριες μέσης τιμή και αυτοσυσχέτισης

Έστω $\{x_i\}_{i=1}^n$ η παρατήρηση μιας στάσιμης χρονοσειράς. Μια εκτιμήτρια για την μέση τιμή δίνεται από την σχέση,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i\,. \tag{220}$$

Μια εκτιμήτρια για την συνάρτηση αυτοσυσχέτισης δίνεται από τη σχέση,

$$\hat{\rho}_\tau = \frac{\hat{\gamma}_\tau}{\hat{\gamma}_0}\,, \tag{221}$$

όπου,

$$\hat{\gamma}_\tau = \frac{1}{n-\tau}\sum_{i=1}^{i-\tau}(x_{t+\tau}-\hat{\mu})(x_t-\hat{\mu})\,. \tag{222}$$

12.4 Εκτιμήτριες των παραμέτρων για το AR μοντέλο

Στην παράγραφο αυτή θα μελετήσουμε διάφορες εκτιμήτριες των παραμέτρων του $AR(p)$ μοντέλου δοσμένης μια παρατήρησης της χρονοσειράς μέχρι την χρόνο T, $\{x_i\}_{i=0}^T$.

12.4.1 Εκτιμήτρια ελαχίστων τετραγώνων

Σύμφωνα με το $AR(p)$ μοντέλο, η παρατήρηση x_n μπορεί να γραφτεί ως,

$$x_n = \sum_{i=1}^{p} a_i x_{n-i} \quad n = p,\ldots,T\,, \tag{223}$$

ή με μορφή πινάκων,

$$\begin{pmatrix} x_p \\ x_{p+1} \\ \vdots \\ x_T \end{pmatrix} = \begin{bmatrix} x_{p-1} & x_{p-2} & \dots & x_0 \\ x_p & x_{p-1} & \dots & x_2 \\ \vdots & \vdots & & \vdots \\ x_{T-1} & x_{T-2} & \dots & x_{T-p} \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}. \tag{224}$$

ή

$$A\boldsymbol{a} = \boldsymbol{x}\,. \tag{225}$$

Επειδή $T > p$ το σύστημα μπορεί να λυθεί με την μέθοδο ελαχίστων τετραγώνων,

$$\boldsymbol{a} = (A^\top A)^{-1} A^\top \boldsymbol{x}\,. \tag{226}$$

12.4.2 Yule-Walker εξισώσεις

Πολλαπλασιάζουμε την σχέση (213) με X_{t-1} και παίρνουμε την μέση τιμή,

$$
\begin{aligned}
\mathbb{E}\left[X_t X_{t-1}\right] &= \sum_{i=1}^{p} a_i \mathbb{E}\left[X_{t-1} X_{t-i}\right] + \mathbb{E}\left[\cancelto{0}{W_t X_{t-1}}\right] \Rightarrow \\
&= \gamma_1 = \sum_{i=1}^{p} a_i \gamma_{i-1} \Rightarrow \\
&= \rho_1 = \sum_{i=1}^{p} a_i \rho_{i-1} \,.
\end{aligned}
\tag{227}
$$

Γενικά, πολλαπλασιάζουμε με $X_j\,, j = 1, \ldots, p$, και παίρνουμε τις Yule-Walker εξισώσεις,

$$
\rho_k = \sum_{i=1}^{p} a_i \rho_{i-k}\,, \quad k = 1, \ldots, p\,.
\tag{228}
$$

Αντικαθιστώντας τις ρ_k από τις εκτιμήτριες $\hat{\rho}_k$ βρίσκουμε τις παραμέτρους a_i.

12.5 Εκτιμήτρια μέγιστης πιθανοφάνειας

Θα δουλέψουμε για την περίπτωση $p = 1$. Παρατηρείστε ότι βάση του μοντέλου $AR(1)$ η τυχαία μεταβλητή X_t εξαρτάται μόνο από την X_{t-1},

$$
X_t \,|\, X_{t-1} \sim \mathcal{N}(a X_{t-1}, \sigma^2)\,.
\tag{229}
$$

Η από κοινού συνάρτηση πυκνότητας πιθανότητας όλης της χρονοσειράς δίνεται από την σχέση,

$$
\begin{aligned}
p(x_T, x_{T-1}, \ldots, x_0) &= p(x_T \,|\, x_{T-1}, \ldots, x_0)\, p(x_{T-1}, \ldots, x_0) \\
&= p(x_T \,|\, x_{T-1})\, p(x_{T-1}, p(x_{T-2}, \ldots, x_0) \\
&= p(x_T \,|\, x_{T-1})\, p(x_{T-1} \,|\, x_{T-2} \ldots, x_0)\, p(x_{T-2}, \ldots, x_0) \\
&= p(x_T \,|\, x_{T-1})\, p(x_{T-1} \,|\, x_{T-2})\, p(x_{T-2}, \ldots, x_0) \\
&\cdots \\
&= p(x_T \,|\, x_{T-1})\, p(x_{T-1} \,|\, x_{T-2}) \cdots p(x_1 \,|\, x_0) \\
&= \prod_{i=1}^{T} p(x_i \,|\, x_{i-1})\,.
\end{aligned}
\tag{230}
$$

Ο λογάριθμος την πιθανοφάνειας υπολογίζεται ως,

$$
\begin{aligned}
\ell(a) &= \sum_{i=1}^{T} \log p(x_i \,|\, x_{i-1}) \\
&= \sum_{i=1}^{T} \left(\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left(x_i - x_{i-1}\right)^2 \right) \\
&= c - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{T} \left(x_i - a x_{i-1}\right)^2
\end{aligned}
\tag{231}
$$

Παίρνοντας την παράγωγο ως προς a και λύνοντας ίση με το μηδέν, παίρνουμε την εκτιμήτρια

$$
\hat{a} = \frac{\hat{\mu}}{\hat{\gamma}_0}\,.
\tag{232}
$$

13 Εβδομάδα 13/13

Επανάληψη