Πυκνοτητα συνεχους τυχαιας μεταβλητης

Συνεχή τυχαία μεταβλητή λέμε οποιαδήποτε τ.μ. της οποίας η συνάρτηση κατανομής $F : \mathbb{R} \to [0,1]$ είναι συνεχής συνάρτηση. Ειδικές περιπτώσεις συνεχών τ.μ. είναι που έχουν πυκνότητα (αυτές λέγονται *απόλυτα συνεχείς*). Δηλαδή αυτές για τις οποίες υπάρχει $f : \mathbb{R} \to [0, \infty)$ Lebesgue-μετρήσιμη με

$$F(b) - F(a) = \int_a^b f(t)\, dt \tag{1}$$

για κάθε $a, b \in \mathbb{R}$ με $a < b$. Μια τέτοια f λέγεται *πυκνότητα* για την X (και για την F), και αν υπάρχει δεν είναι μοναδική. Αλλάζοντας την σε πεπερασμένα το πλήθος σημεία εξακολουθεί να ικανοποιεί την παραπάνω ισότητα. Αποδεικνύεται ότι πυκνότητες είναι ακριβώς οι (Lebesgue-μετρήσιμες) συναρτήσεις που ικανοποιούν $f \geqslant 0, \int_{\mathbb{R}} f(x)\, dx = 1$.

Ερώτημα: Δεδομένης μιας συνεχούς συνάρτησης κατανομής F, πώς αποδεικνύουμε ότι αυτή προέρχεται από πυκνότητα, και πώς βρίσκουμε μια πυκνότητα f;

Αν η F προέρχεται απο πυκνότητα, τότε θα έχουμε $f(x) = F'(x)$ σε πάρα πολλά σημεία (σε όλα εκτός απο ένα σύνολο μέτρου μηδέν). Αυτό είναι ένα σημείο εκκίνησης.

Για το μάθημα, θα μας αρκέσει η εξής πρόταση. (Για πιο προχωρημένα αποτελέσματα, δες στην τελευταία σελίδα αυτού του φυλλαδίου.)

Πρόταση. Έστω συνεχής συνάρτηση κατανομής F για την οποία υπάρχει ένα πεπερασμένο σύνολο $J = \{a_1, a_2, \ldots, a_n\} \subset \mathbb{R}$ και συνάρτηση $f : \mathbb{R} \to [0, \infty)$ έτσι ώστε

- $F'(t) = f(t)$ για κάθε $t \in \mathbb{R} \setminus \{a_1, a_2, \ldots, a_n\}$,
- f συνεχής στο $\mathbb{R} \setminus \{a_1, a_2, \ldots, a_n\}$.

Τότε η f είναι μια πυκνότητα που αντιστοιχεί στη συνάρτηση κατανομής F.

Συμπέρασμα: Δεδομένης μιας συνεχούς F, υπολογίζουμε την F', και έστω ότι αυτή υπάρχει παντού εκτός από ένα πεπερασμένο σύνολο J_1. Το ευνοϊκό σενάριο είναι η F' να είναι συνεχής στο $\mathbb{R} \setminus J_1$, ή αν είναι ασυνεχής, να είναι ασυνεχής σε πεπερασμένο σύνολο σημείων J_2 του $\mathbb{R} \setminus J_1$. Θέτουμε τότε $J := J_1 \cup J_2$ και $f := F'$ στο $\mathbb{R} \setminus J$. Επεκτείνουμε την f αυθαίρετα στο J δίνοντάς της κάποιες μη αρνητικές τιμές. Για παράδειγμα, $f = 0$ στο J. Τότε για τις F, f εφαρμόζεται η πιο πάνω πρόταση και παίρνουμε ότι η f είναι μια πυκνότητα για την F.

Εφαρμογη

Έστω X συνεχής τυχαία μεταβλητή με πυκνότητα

$$f_X(x) = \frac{1}{\pi} \frac{1}{(1 + x^2)} \text{ για κάθε } x \in \mathbb{R}.$$

Θέτουμε $Y = \tan^{-1} X$, όπου $\tan^{-1}$ είναι η αντίστροφη του περιορισμού της $\tan$ στο $(-\pi/2, \pi/2)$. Ποια είναι η κατανομή της Y;

Λύση

Για $t \in \mathbb{R}$ έχουμε $F_Y(t) := P(Y \leqslant t)$, το οποίο ισούται με 0 για $t \leqslant -\pi/2$ και με 1 για $t \geqslant \pi/2$ επειδή η Y παίρνει τιμές στο $(-\pi/2, \pi/2)$. Έπειτα, για $t \in (-\pi/2, \pi/2)$,

$$F_Y(t) = P(\tan^{-1} X \leqslant t) = P(X \leqslant \tan t) = F_X(\tan t).$$

Άρα

$$F_Y(t) = \begin{cases} 0 & \text{αν } t \leqslant -\pi/2, \\ F_X(\tan t) & \text{αν } t \in (-\pi/2, \pi/2), \\ 1 & \text{αν } t \leqslant -\pi/2. \end{cases}$$

Επειδή $F_X(-\infty) = 0$ και $F_X(\infty) = 1$, έπεται ότι η F_Y είναι συνεχής. Επίσης, για $t \in (-\pi/2, \pi/2)$,

$$F'_Y(t) = F'_X(\tan t)(\tan t)' = \frac{1}{\pi(1+\tan^2 t)}(1+\tan^2 t) = \frac{1}{\pi}.$$

Και βέβαια $F'_Y(t) = 0$ για $|t| > \pi/2$. Επομένως η F'_Y υπάρχει και είναι συνεχής στο συμπλήρωμα του πεπερασμένου συνόλου $\{-\pi/2, \pi/2\}$. Με βαση την παραπάνω πρόταση, η Y είναι συνεχής με πυκνότητα, και μια πυκνότητα για την Y είναι η

$$f_Y(t) := \begin{cases} 0 & \text{αν } |t| \geqslant \pi/2, \\ \frac{1}{\pi} & \text{αν } t \in (-\pi/2, \pi/2). \end{cases}$$

Άρα η Y έχει την ομοιόμορφη κατανομή στο $(-\pi/2, \pi/2)$.

Αποδειξη της προτασης

Η απόδειξη είναι άμεση συνέπεια του δεύτερου θεμελιώδους θεωρήματος του ολοκληρωτικού λογισμού και του ορισμού του γενικευμένου ολοκληρώματος. Πάει ως εξής.
Θέλουμε να δείξουμε ότι για κάθε $a < b$ ισχύει

$$F(b) - F(a) = \int_a^b f(t)\,dt. \tag{2}$$

Αν $J = \emptyset$, τότε η (2) έπεται από το δεύτερο θεμελιώδες θεώρημα του ολοκληρωτικού λογισμού ($f = F'$ και η F' είναι συνεχής στο $[a, b]$, και άρα ολοκληρώσιμη). Αν $J \neq \emptyset$, διακρίνουμε τις εξής περιπτώσεις.

Περίπτωση 1. Αν το μόνο σημείο του J στο $[a, b]$ είναι το a, τότε επειδή η F είναι συνεχής έχουμε

$$F(b) - F(a) = \lim_{\varepsilon \to 0^+} \{F(b) - F(a+\varepsilon)\}.$$

Επειδή η F' είναι συνεχής στο $[a+\varepsilon, b]$, το δεύτερο θεμελιώδες θεώρημα του ολοκληρωτικού λογισμού δίνει

$$F(b) - F(a+\varepsilon) = \int_{a+\varepsilon}^b F'(t)\,dt = \int_{a+\varepsilon}^b f(t)\,dt.$$

Το όριο του τελευταίου ολοκληρώματος για $\varepsilon \to 0^+$ υπάρχει γιατί είναι φθίνουσα συνάρτηση του ε (αφού $f \geqslant 0$), και είναι εξ ορισμού η τιμή του $\int_a^b f(t)\,dt$. Για παράδειγμα, είναι δυνατόν να έχουμε $f(a+) = \infty$ (όπως συμβαίνει με την $1/\sqrt{x}$ στο $a = 0$) ή το δεξιό όριο της f στο a να μην υπάρχει.

Περίπτωση 2. Αν το μόνο σημείο του J στο $[a, b]$ είναι το b. Δουλεύουμε όμοια όπως στην προηγούμενη περίπτωση. Πάλι η (2) ισχύει.

Περίπτωση 3. Αν τα μόνα σημεία του J του στο $[a, b]$ είναι τα a, b. Επιλέγουμε $c \in (a, b)$, εφαρμόζουμε την (2) στα διαστήματα $[a, c], [c, b]$ (το οποίο ξέρουμε από τις προηγούμενες περιπτώσεις), και προσθέτουμε κατά μέλη.

Περίπτωση 4. Αν τα σημεία του J ανάμεσα στα a, b είναι τα $x_1, x_2, \ldots, x_k$ με $a \leqslant x_1 < x_2 < \cdots < x_k \leqslant b$, τότε εφαρμόζουμε την (2) στα διαστήματα $[a, x_1], [x_1, x_2], \ldots, [x_k, b]$, και προσθέτουμε τις ισότητες που προκύπτουν κατά μέλη. ■

Είναι προφανές απο την απόδειξη της πρότασης ότι η υπόθεση «J πεπερασμένο» μπορεί να αντικατασταθεί απο την

$$J \cap [a,b] \text{ πεπερασμένο για κάθε πεπερασμένο υποδιάστημα } [a,b] \text{ του } \mathbb{R},$$

δηλαδή το J δεν έχει σημεία συσσώρευσης στο $\mathbb{R}$.

Πιο προχωρημενα αποτελεσματα

Επιστρέφουμε στο ερώτημα που τέθηκε στην αρχή αυτή της παραγράφου. Δηλαδή, αν δίνεται η συνάρτηση κατανομής, F, μιας τυχαίας μεταβλητής, μπορούμε να αποφανθούμε αν αυτή η τυχαία μεταβλητή έχει πυκνότητα; Η ικανή και αναγκαία συνθήκη είναι η F να είναι **απόλυτα συνεχής** (δες Θεώρημα 14.22 στο [1]). Δηλαδή για κάθε $\varepsilon > 0$ υπάρχει $\delta > 0$ ώστε αν το $n \in \mathbb{N}^+$ και $\{(a_i, b_i) : i = 1, 2, \ldots, n\}$ είναι ξένα ανά δύο διαστήματα με $\sum_{i=1}^n (b_i - a_i) < \delta$, τότε ισχύει

$$\sum_{i=1}^{n} |F(b_i) - F(a_i)| < \varepsilon. \tag{3}$$

Αυτό ισχύει, για παράδειγμα, αν η F είναι Lipschitz.

Ικανές συνθήκες ώστε να είναι η F (και γενικότερα μια αύξουσα F) απόλυτα συνεχής δίνουν τα ακόλουθα αποτελέσματα. Το δεύτερο είναι γενίκευση του πρώτου και έχει πιο δύσκολη απόδειξη.

Πρόταση 1. Αν η F είναι συνεχής (στο $\mathbb{R}$) και υπάρχει σύνολο $A \subset \mathbb{R}$ χωρίς σημείο συσσώρευσης στο $\mathbb{R}$ ώστε η F' να υπάρχει και να είναι πεπερασμένη για κάθε $x \in \mathbb{R} \setminus A$, τότε η F είναι απόλυτα συνεχής.

Πρόταση 2. Αν η F είναι συνεχής (στο $\mathbb{R}$) και υπάρχει σύνολο $A \subset \mathbb{R}$ αριθμήσιμο ώστε η F' να υπάρχει και να είναι πεπερασμένη για κάθε $x \in \mathbb{R} \setminus A$, τότε η F είναι απόλυτα συνεχής.

Και στις δύο περιπτώσεις μια πυκνότητα είναι η

$$f(x) := \begin{cases} F'(x) & \text{αν } x \in \mathbb{R} \setminus A, \\ 0 & \text{αν } x \in A. \end{cases}$$

Η απόδειξη της Πρότασης 1 έπεται από το Θεώρημα 7.19 και το Πόρισμα 14.9 στο [1] ενώ της 2 από το Πόρισμα 14.9 και την Άσκηση 14-15 στο [1]. Στην ξένη βιβλιογραφία, το θέμα των απολύτως συνεχών συναρτήσεων καλύπτεται, για παράδειγμα, στο Κεφάλαιο 7 του [3] και στην Παράγραφο 18 του [2].

[1] Γ. Κουμουλλής, Σ. Νεγρεπόντης. Θεωρία Μέτρου. Εκδόσεις Συμμετρία.
[2] E. Hewitt-K. Stromberg. Real and abstract analysis. Springer-Verlag, 1969.
[3] R. Kannan-C.K. Krueger. Advanced analysis on the real line. Springer-Verlag, 1996.