Συνάρτηση Πυκνότητας

Ορισμός: Έστω κατανομή πιθανότητας στον $\mathbb{R}$ με αθροιστική συνάρτηση $F : \mathbb{R} \to [0, 1]$

Εάν υπάρχει f τέτοια ώστε $F(x) = \int_{-\infty}^{x} f(z)\,dz$ τότε η f ονομάζεται συνάρτηση πυκνότητας της κατανομής (probability density function - pdf).

Θεώρημα Ύπαρξης

προκειμένου να υπάρχει η f είναι απαραίτητο

α. η F να είναι παντού συνεχής

β. η F να είναι σχεδόν παντού* παραγωγίσιμη

* επιτρέπεται να υπάρχουν x στα οποία η F δεν είναι παραγωγίσιμη, αρκεί να είναι απομονωμένα μεταξύ τους (δηλαδή να συγκροτούν ένα διακριτό σύνολο)

Συνεπώς, οι διακριτές κατανομές δεν έχουν συνάρτηση πυκνότητας (π.χ. εκφυλισμένη, Bernoulli, Poisson, μεικτή, συνεχής P με ασυνεχή F)

Πως εξάγουμε την f;

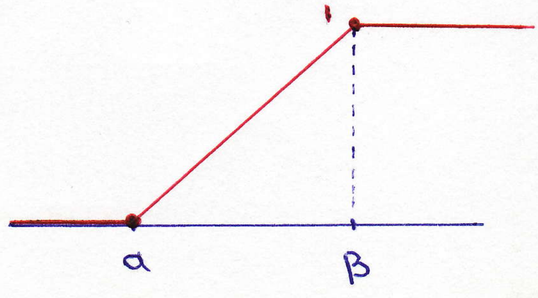
- όπου η F είναι παραγωγίσιμη την παραγωγίζουμε ως προς x, $f = \frac{dF}{dx}$
- στα σημεία όπου η F δεν παραγωγίζεται, δίνουμε στην f αυθαίρετες τιμές. Για λόγους συνεννόησης επιλέγουμε τις τιμές αυτές για τις οποίες η f είναι συνεχής στο στήριγμα

Παράδειγμα: unif$[\alpha, \beta]$ | Εφαρμόζουμε το Θεώρημα Ύπαρξης

α. γνωρίζουμε ότι η ομοιόμορφη (uniform) κατανομή P είναι συνεχής, διότι το στήριγμά της είναι συνεχές, $\text{supp}\, P_{unif} = [\alpha, \beta]$

Συνεπώς, η $F = \begin{cases} 0, & x < \alpha \\ \dfrac{x-\alpha}{\beta-\alpha}, & \alpha \leq x < \beta \\ 1, & x \geq \beta \end{cases}$

είναι παντού συνεχής



Β. σχεδόν παντού παραγωγίσιμη F (εκτός α, β)

παραγωγίζουμε όπου η F είναι παραγωγίσιμη

- $x < \alpha \rightarrow f(x) = \frac{dF}{dx} = \frac{d(0)}{dx} = 0$ σταθερή
- $\alpha < x < \beta \rightarrow f(x) = \frac{dF}{dx} = \frac{d}{dx}\left(\frac{x-\alpha}{\beta-\alpha}\right) = \frac{1}{\beta-\alpha}$ σταθερή
- $x > \beta \rightarrow f(x) = \frac{dF}{dx} = \frac{d(1)}{dx} = 0$ σταθερή

Στα διαστήματα εντός και εκτός supp η F είναι παραγωγίσιμη. Στα α, β δεν είναι (δεν μπορούμε να φέρουμε εφαπτομένη με μοναδικό τρόπο). Συνεπώς, μπορούμε να δώσουμε αυθαίρετες τιμές, έστω $f(\alpha) = c_1$, $f(\beta) = c_2$, όπου $c_1, c_2 \in \mathbb{R}$.

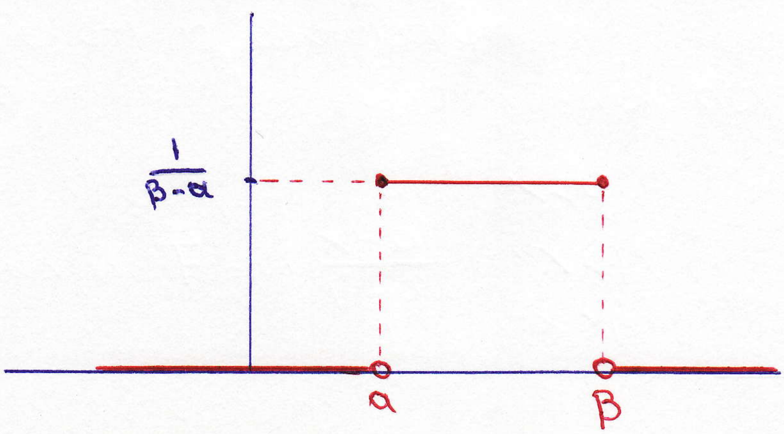
$$f(x) = \begin{cases} 0, & x < \alpha \\ c_1, & x = \alpha \\ \frac{1}{\beta-\alpha}, & \alpha < x < \beta \\ c_2, & x = \beta \\ 0, & x > \beta \end{cases}$$

Ανάλογα με την επιλογή των c_1, c_2 υπάρχουν άπειρες $f(x)$. Συνεπώς, η f δεν είναι μοναδική. Αυτό όμως δεν επηρεάζει τις πιθανότητες διότι τα α, β είναι απομονωμένα. Συμβατικά, διαλέγουμε τιμές τέτοιες ώστε η f να είναι συνεχής στο στήριγμα.

Εδώ, supp $P_{unif} = [\alpha, \beta]$. Θέτουμε $f(\alpha) = f(\beta) = \frac{1}{\beta-\alpha}$

Άρα, $f(x) = \begin{cases} 0, & x < \alpha \\ \frac{1}{\beta-\alpha}, & \alpha \le x \le \beta \\ 0, & x > \beta \end{cases}$

Εφόσον, ικανοποιείται το Θεώρημα Ύπαρξης (α & β) η f_{unif} ορίζεται.



Θεώρημα Χαρακτηρισμού

Η συνάρτηση πυκνότητας f θα είναι καλώς ορισμένη αν ισχύουν

α. $f(x) \geq 0, \ \forall x \in \mathbb{R}$

β. $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

όπου $f(x) = \frac{dF(x)}{dx}$ στα σημεία παραγωγισιμότητας της F, ενώ στα σημεία μη παραγωγισιμότητας της F, η f παίρνει αυθαίρετες τιμές.

Συνεπώς, προκειμένου να αποδείξουμε αν οποιαδήποτε συνάρτηση f είναι pdf, ελέγχουμε την εγκυρότητα των α, β.

Ανακεφαλαιώνοντας

1. Η pdf δεν ορίζεται για κάθε κατανομή, σε αντίθεση με την cdf που ορίζεται πάντα.
2. Αν ορίζεται, μπορεί να μην είναι μοναδική (αυτό συμβαίνει όταν η F έχει σημεία μη παραγωγισιμότητας), σε αντίθεση με την F που είναι πάντα μοναδική.
3. Βάσει του θεωρήματος χαρακτηρισμού, μπορούμε να χρησιμοποιήσουμε την f για να υπολογίσουμε πιθανότητες.
4. Επιλέγουμε την f που είναι συνεχής στο στήριγμα.

Άσκηση 1

1. Νδο η παρακάτω συνάρτηση είναι pdf
2. Να βρεθεί η πιθανότητα $P(X \leq 4)$
3. Να βρεθεί η αθροιστική συνάρτηση F της

$$f(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{x^2}, & x \geq 1 \end{cases}$$

Λύση

1. Θεώρημα Χαρακτηρισμού

α. $f(x) \geq 0, \ \forall x \in \mathbb{R}$ πράγματι ισχύει

β. $\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{1} 0\, dx + \int_{1}^{+\infty} \frac{1}{x^2} dx = 0 + \left[-\frac{1}{x}\right]_1^{+\infty} = \lim_{x \to +\infty}\left(-\frac{1}{x}\right) + \frac{1}{1} = 1$ ✓

2. $P(X \leq 4) = P((-\infty, 4]) = \int_{-\infty}^{4} f(x)\,dx = \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{4} f(x)\,dx$

$= 0 + \int_{1}^{4} \frac{1}{x^2}\,dx = \left[-\frac{1}{x}\right]_1^4 = -\frac{1}{4} + 1 = \frac{3}{4}$

3. Εξ ορισμού γνωρίζουμε ότι $F(x) = \int_{-\infty}^{x} f(z)\,dz$

- $x < 1$, $F(x) = \int_{-\infty}^{x} f(z)\,dz = \int_{-\infty}^{x} 0\,dz = 0$
- $x \geq 1$, $F(x) = \int_{-\infty}^{x} f(z)\,dz = \int_{-\infty}^{1} 0\,dz + \int_{1}^{x} \frac{1}{z^2}\,dz$

$= 0 + \left[-\frac{1}{z}\right]_1^x = -\frac{1}{x} + \frac{1}{1} = 1 - \frac{1}{x}$

Η $F(x) = \begin{cases} 0, & x < 1 \\ 1 - \frac{1}{x}, & x \geq 1 \end{cases}$

Υπολογισμός Πιθανοτήτων με την f

Ορισμός: Αν η $f(x)$ είναι η pdf της X, τότε η πιθανότητα το x να ανήκει σε ένα διάστημα A, δίνεται από το ολοκλήρωμα της $f(x)$ σε αυτό το διάστημα:

$$P(X \in A) = \int_A f(x)\,dx$$

π.χ. Έστω X τυχαία μεταβλητή της οποίας η pdf είναι η $f(x) = 3x^2$, $0 < x < 1$.

Αρχικά, παρατηρούμε ότι $f(x) \neq P(X = x)$.
$x = 0.9 \Rightarrow f(0.9) = 2.43$. Στην περίπτωση των συνεχών κατανομών η $f(x)$ είναι το ύψος της καμπύλης όπου $X = x$, έτσι ώστε όλη η περιοχή κάτω από την καμπύλη να είναι 1.
Είναι οι περιοχές κάτω από την καμπύλη που ορίζουν πιθανότητες.

$$P\left(\frac{1}{2} < X < 1\right) = \int_{\frac{1}{2}}^{1} 3x^2\,dx = \left[\frac{x^3}{3}\right]_{\frac{1}{2}}^{1} 3 = \frac{7}{8}$$

- $P(X=\frac{1}{2}) = \int_{\frac{1}{2}}^{\frac{1}{2}} 3x^2\,dx = [x^3]_{\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{8}-\frac{1}{8}=0$

Γενικά, αν η X είναι συνεχής, η πιθανότητα σημείου είναι 0, $P(X=x)=0,\ \forall\, x\in \text{supp}X$. Αυτό σημαίνει ότι τα άκρα του ολοκληρώματος δεν επηρεάζουν τον υπολογισμό πιθανοτήτων συνεχών τυχαίων μεταβλητών.

$$P(a\le X\le \beta)=P(a<X\le\beta)=P(a\le X<\beta)=P(a<X<\beta)$$

Άσκηση 2

Δίνεται η $f(x)=\begin{cases} cx^2, & 0<x<3 \\ 0, & \text{αλλιώς}\end{cases}$

α. Να βρεθεί το c έτσι ώστε η $f(x)$ να είναι pdf

β. Να βρεθεί η αθροιστική συνάρτηση F

Λύση

α. Θεώρημα Χαρακτηρισμού

- $f(x)\ge 0 \Rightarrow cx^2\ge 0 \Rightarrow c\ge 0$
- $\int_{-\infty}^{+\infty} f(x)\,dx=1 \Rightarrow \int_{-\infty}^{0} 0\,dx+\int_{0}^{3} cx^2dx+\int_{3}^{+\infty} 0\,dx=1 \Leftrightarrow c\left[\frac{x^3}{3}\right]_0^3=1$

$\Leftrightarrow c\left\{\frac{27}{3}-\frac{0}{3}\right\}=1 \Leftrightarrow 9c=1 \Rightarrow c=\frac{1}{9}$ $\Big|\ f(x)=\begin{cases} \frac{1}{9}x^2, & 0<x<3 \\ 0, & \text{αλλιώς}\end{cases}$

β. Εξ ορισμού γνωρίζουμε ότι $F(x)=\int_{-\infty}^{x} f(z)\,dz$

- $x<0 \Rightarrow F(x)=\int_{-\infty}^{x} 0\,dz=0$
- $0\le x<3 \Rightarrow F(x)=\int_{-\infty}^{x} f(z)dz=\int_{-\infty}^{0} 0\,dz+\int_{0}^{x}\frac{1}{9}z^2dz=\left[\frac{z^3}{3}\right]_0^x\cdot\frac{1}{9}=\frac{x^3}{27}$
- $x\ge 3 \Rightarrow F(x)=\int_{-\infty}^{x} f(z)dz=\int_{-\infty}^{0} 0\,dz+\int_{0}^{3}\frac{1}{9}z^2dz+\int_{3}^{x} 0\,dz=$

$=\frac{1}{9}\left[\frac{z^3}{3}\right]_0^3=\frac{1}{9}\left[\frac{27}{3}-\frac{0}{3}\right]=\frac{27}{27}=1$

Συνεπώς $F(x)=\begin{cases} 0, & x<0 \\ \frac{x^3}{27}, & 0\le x<3 \\ 1, & x\ge 3\end{cases}$