Κατανομές Πιθανότητας στους Πραγματικούς- Αθροιστική Συνάρτηση

Η ευκολία στην περιγραφή των διακριτών κατανομών δεν ισχύει γενικότερα. Μας χρειάζονται έννοιες, μέσω των οποίων να μπορούμε να περιγράφουμε μια κατανομή και οι οποίες μάλιστα να μας είναι πιο "οικείες", π.χ. συναρτήσεις $\mathbb{R}\to\mathbb{R}$. Προφανώς, εφόσον τέτοιες υπάρχουν, έχει ενδιαφέρον το πως οι ιδιότητες των τελευταίων αναπαριστούν ιδιότητες των κατανομών.

Η πρώτη έννοια που συναντούμε είναι αυτή της αθροιστικής συνάρτησης.

Ορισμός. Έστω P κατανομή στο $\mathbb{R}$. Η αθροιστική συνάρτηση της κατανομής (cdf – cumulative distribution function), είναι η $F:\mathbb{R}\to\mathbb{R}$, που ορίζεται από:

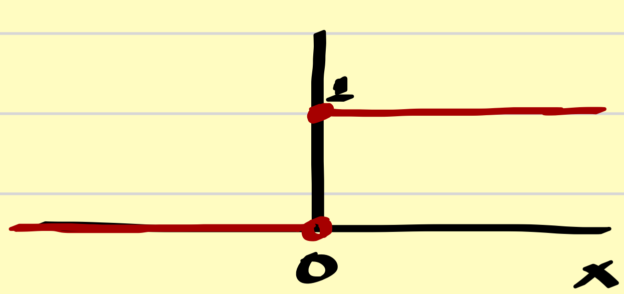
$$F(x) = P((-\infty,x]), \quad \forall x\in\mathbb{R}. \square$$

Θυμηθείτε ότι το $(-\infty,x]\in\Sigma_{\mathbb{R}}$ ως διάστημα. Επομένως η $P((-\infty,x])$ υπάρχει $\forall x\in\mathbb{R}$. Συνεπώς η F είναι καλώς ορισμένη συνάρτηση $\mathbb{R}\to\mathbb{R}$.

Παράδειγμα. Έστω η εκφυλισμένη κατανομή στο 0.
Έχουμε ότι

$$F(x)=P((-\infty,x])=P((-\infty,x]\cap\{0\})=\begin{cases}P(\emptyset), & x<0\\ P(\{0\}), & x\geq 0\end{cases}$$

$$=\begin{cases}0 & x<0\\ 1 & x\geq 0,\end{cases}$$ με γράφημα

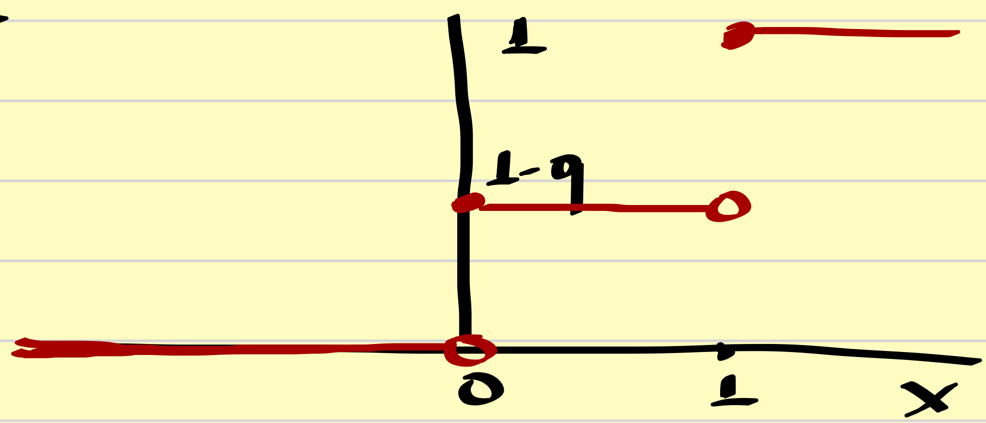


$\square$

Παράδειγμα. Έστω η Ber(q). Έχουμε ότι

$$F(x) = P((-\infty,x]) = P((-\infty,x]\cap\{0,1\})$$
$$= \begin{cases} P(\emptyset), & x<0 \\ P(\{0\}), & 0\le x<1 \\ P(\{0,1\}), & x\ge 1 \end{cases} = \begin{cases} 0, & x<0 \\ 1-q, & 0\le x<1 \\ 1, & x\ge 1 \end{cases}$$

με γράφημα



□

Παρατηρούμε ότι και στα δύο παραδείγματα οι cdf είναι «φραγμένες», αύξουσες, μπορεί να είναι ασυνεχείς σε σημεία, αλλά είναι παντού από δεξιά συνεχείς.

Θεώρημα. Η F έχει τις εξής τρεις ιδιότητες:

1. $\lim_{x\to-\infty} F(x)=0$, $\lim_{x\to+\infty} F(x)=1$
2. η F είναι αύξουσα, δηλ. αν $x_1<x_2 \Rightarrow F(x_1)\le F(x_2)$,
3. η F είναι από δεξιά συνεχής, δηλ. $\forall x\in\mathbb{R}$, $\lim_{y\to x^+} F(y) = F(x)$.

Αντίστροφα αν κάποια $F:\mathbb{R}\to\mathbb{R}$ έχει τις ιδιότητες 1,2,3 τότε είναι η cdf μοναδικής κατανομής P στο $\mathbb{R}$. □

Σχόλια: Το παραπάνω είναι θεώρημα αναπαράστασης, αφού δεδομένου ότι σε κάθε P αντιστοιχεί μοναδική

P, αυτό μας λέει ότι ισχύει και το αντίστροφο. Δηλαδή και επειδή οι 1,2,3 αποτελούν χαρακτηριστικές ιδιότητες ενός cdf, σε κάθε $F:\mathbb{R}\to\mathbb{R}$ που τις έχει αντιστοιχεί μοναδική P στο $\mathbb{R}$. Αυτό σημαίνει ότι η F αναπαριστά "τέλεια" την P, οπότε μπορούμε να ορίζουμε την P μέσω της F, να βρίσκουμε τις πιθανότητες που αυτή αποδίδει, ιδιότητες της P θα αντανακλώνται σε ιδιότητες της F κ.ο.κ. Κάποια από αυτά τα διερευνούμε παρακάτω.

Θα σκιαγραφήσουμε την απόδειξη των 1,2 και 3. Οι συλλογισμοί που θα παρουσιαστούν δεν θα είναι πλήρεις καθώς οι αποδείξεις βασίζονται σε κάποια ιδιότητα "συνέχειας" της P που δεν έχουμε αναλύσει. Επομένως όπου χρησιμοποιούνται στα παρακάτω αποσιωπητικά θα εννοείται η χρήση αυτής της ιδιότητας. Επίσης η απόδειξη του τελευταίου μέρους του παραπάνω θεωρήματος είναι εκτός του εύρους του μαθήματος ενώ βασίζεται στο ότι τα διαστήματα της μορφής $(-\infty,x]$, $\forall x\in\mathbb{R}$ κατά κάποιο τρόπο παρέχουν αρκετή πληροφορία για το $\Sigma_{\mathbb{R}}$.

Σκιαγράφηση της απόδειξης των ιδιοτήτων 1,2,3.

1. $\lim\limits_{x\to-\infty} F(x) = \lim\limits_{x\to-\infty} P(-\infty,x] = \ldots = P\left(\bigcap_{n=1}^{\infty}(-\infty,-n]\right)$ $(*)$

$\bigcap_{n=1}^{\infty}(-\infty,-n] = (-\infty,-1]\cap(-\infty,-2]\cap\ldots\cap(-\infty,-n]\cap\ldots$

$=\emptyset$ αφού $\forall x\in\mathbb{R}, \exists n\in\mathbb{N}: -n<x \Rightarrow x\notin(-\infty,-n]$.
Επομένως $(*) = P(\emptyset)=0$.

$$\lim_{x\to+\infty} F(x) = \lim_{x\to+\infty} P(-\infty,x] = \ldots = P\left(\bigcup_{n=1}^{\infty}(-\infty,n]\right) \quad (**)$$

$\bigcup_{n=1}^{\infty}(-\infty,n] = (-\infty,1]\cup(-\infty,2]\cup\ldots\cup(-\infty,n]\cup\ldots = \mathbb{R}$ αφού

$\forall x \in \mathbb{R}, \exists n \in \mathbb{N}: x \leq n \Rightarrow x \in (-\infty, n]$. Επομένως
$(**) = P(\mathbb{R}) = 1$.

2. Έστω $x_1 < x_2 \Rightarrow (-\infty, x_1] \subset (-\infty, x_2] \underset{\text{μονοτονία της } P}{\Rightarrow} P((-\infty, x_1]) \leq P((-\infty, x_2])$

$\Leftrightarrow F(x_1) \leq F(x_2)$. (Παρατηρούμε ότι η μονοτονία της P αντανακλάται ευθέως σε μονοτονία της F, κάτι το οποίο αποτελεί πιο "οικεία" ιδιότητα αφού η F είναι συνάρτηση από το $\mathbb{R}$ στο $\mathbb{R}$).

3. Έστω $x \in \mathbb{R}$. $\lim_{y \to x^+} F(y) = \lim_{y \to x^+} P((-\infty, y]) = \ldots =$

$$= P\left(\bigcap_{n=1}^{\infty} (-\infty, x + \tfrac{1}{n}]\right) \quad (***)$$

$$\bigcap_{n=1}^{\infty} (-\infty, x + 1/n] = (-\infty, x+1] \cap (-\infty, x + 1/2] \cap \ldots \cap (-\infty, x + \tfrac{1}{n}] \cap \ldots$$

Προφανώς το $(-\infty, x]$ είναι υποσύνολο της παραπάνω τομής (γιατί;), αλλά αν $z > x \Rightarrow \exists n \in \mathbb{N}: z > x + \frac{1}{n} \Rightarrow z \notin (-\infty, x + \frac{1}{n}] \Rightarrow$ $z \notin \bigcap_{n=1}^{\infty} (-\infty, x + \frac{1}{n}]$. Επομένως $\bigcap_{n=1}^{\infty} (-\infty, x + \frac{1}{n}] = (-\infty, x]$ και

$(***) = P((-\infty, x]) = F(x)$. $\square$

Από τις παραπάνω ιδιότητες προκύπτουν και περαιτέρω ιδιότητες της cdf κάποιες από τις οποίες περιγράφονται στα παρακάτω.

Έχουμε ότι αν $x \in \mathbb{R}$, τότε $\lim_{y \to x^-} F(y) = \lim_{y \to x^-} P((-\infty, y]) = \ldots =$

$$= P\left(\bigcup_{n=1}^{\infty} (-\infty, x - \tfrac{1}{n}]\right) \underset{[\Delta]}{=} P((-\infty, x)).$$

Άσκηση – Δείξε το $[\Delta]$.

Η F θα είναι συνεχής στο x εξαιτίας των 2,3 αν

$F(x) > \lim_{y\to x^-} F(y)$ (γιατί;) $\iff F(x) - \lim_{y\to x^-} F(y) > 0 \iff$ (βλ. προηγούμενο) $P((-\infty,x]) - P((-\infty,x)) > 0 \overset{\text{γιατί;}}{\iff} P(\{x\})>0$.

Επομένως αποδείξαμε το εξής αποτέλεσμα.

Πόρισμα. Η F ασυνεχής στο x ανν $P(\{x\})>0$ ($\iff$ η F συνεχής στο x ανν $P(\{x\})=0$). $\square$

Επομένως η F οποιασδήποτε διακριτής κατανομής θα εμφανίζει ασυνέχεια σε κάθε στοιχείο του στηρίγματος, και μόνο εκεί, κάτι το οποίο εξηγεί την ανάλογη συμπεριφορά στα δύο παραπάνω παραδείγματα.

Επιπλέον, έστω $(x_1,x_2] \subseteq supp'$. Έχουμε ότι $F(x_2) - F(x_1) = P((-\infty,x_2]) - P((-\infty,x_1]) = P((x_1,x_2]) \leq P(supp') = 0$. Επομένως (γιατί;) $F(x_1) = F(x_2) \overset{2}{\Longrightarrow}$ η F σταθερή στο $(x_1,x_2]$. Αυτό συνεπάγεται το παρακάτω:

Πόρισμα. Η F σταθερή σε διάστημα ανν το διάστημα είναι κομμάτι του $supp'$ ($\iff$ η F γνησίως αύξουσα σε διαστήματα που είναι κομμάτι του $supp$ (γιατί;)).

Υπολογισμός Πιθανοτήτων Βάσει της cdf

Εφόσον η F αναπαριστά την P θα πρέπει μέσω της F να μπορούμε να υπολογίζουμε τις πιθανότητες που η P αποδίδει. Μερικά, προφανώς μη εξαντλητικά, παραδείγματα είναι τα εξής: αν $\alpha<\beta$

$$P((-\infty,\alpha]) = F(\alpha),$$ από τον ορισμό

$P((-\infty,\alpha)) = \lim_{y\to\alpha^-} F(y)$, από προηγουμένως

$P(\{\alpha\}) = F(\alpha) - \lim_{y\to\alpha^-} F(y)$, από προηγουμένως

$P((\alpha,\beta]) = F(\beta) - F(\alpha)$, από προηγουμένως - αυτό θυμίζει το θεμελιώδες θεώρημα του λογισμού.

$$P((\alpha,\beta)) = P((-\infty,\beta)) - P((-\infty,\alpha]) = \lim_{y\to\beta^-} F(y) - F(\alpha),$$

$$P([\alpha,\beta]) = P((-\infty,\beta]) - P((-\infty,\alpha)) = F(\beta) - \lim_{y\to\alpha^-} F(y),$$

$$P([\alpha,\beta)) = P((-\infty,\beta)) - P((-\infty,\alpha)) = \lim_{y\to\beta^-} F(y) - \lim_{y\to\alpha^-} F(y),$$

Όταν τα α,β είναι σημεία συνέχειας της F, τότε $P([\alpha,\beta]) = P((\alpha,\beta)) = P([\alpha,\beta)) = P((\alpha,\beta])$ - γιατί;

$$P((\alpha,+\infty)) = 1 - P((-\infty,\alpha]) = 1 - F(\alpha),$$

$$P([\alpha,+\infty)) = 1 - P((-\infty,\alpha)) = 1 - \lim_{y\to\alpha^-} F(y), \text{ κ.ο.κ.}$$

Οι σημειώσεις βρίσκονται σε στάδιο διαρκούς διόρθωσης και δεν υποκαθιστούν τις διαλέξεις. Παρακαλώ αναφέρτε όποιο λάθος στο stelios@aueb.gr ή στο e-class του μαθήματος.