Διάλεξη 19

Στοιχεία της θεωρίας της συνάρτησης πυκνότητας και παραδείγματα

B. Συνάρτηση Πυκνότητας (Density function)

Στο παράδειγμα της $N(\mu,v)$ είδαμε ότι η αθροιστική έχει την μορφή ορισμένου ολοκληρώματος κάποιας άλλης συνάρτησης. Γιατί μπορεί να είναι χρήσιμο κάτι τέτοιο;

Τουλάχιστον για δύο λόγους:
- Επιβοηθητικό στο να ανατηλίσουμε τις κατανομές ως διαδικασίες ολοκλήρωσης.
- Σχετίζεται με αυτό που ονομάζεται συνάρτηση πιθανοφάνειας στην στατιστική επαγωγή.

Ορισμός. Έστω $\mathbb{P}$ κατανομή στο $\mathbb{R}$ και F η αθροιστική της. Αν υπάρχει $f:\mathbb{R}\to\mathbb{R}$ τέτοια ώστε να έχουμε ότι

$$F(x) = \int_{-\infty}^{x} f(z)\,dz, \quad \forall x\in\mathbb{R}, \text{ τότε η } f$$

ονομάζεται συνάρτηση πυκνότητας της $\mathbb{P}$ (probability density function-pdf).

Πότε υπάρχει η συνάρτηση πυκνότητας;

Είναι δυνατόν να αποδειχθεί ότι αναγκαία και ικανή συνθήκη για να υπάρχει η f, η F να ικανοποιεί κάποια ιδιότητα συνέχειας που είναι ισχυρότερη της συνηθισμένης (ονομάζεται απόλυτη συνέχεια - absolute continuity και είναι εκτός του εύρους του μαθήματος)

Έτσι λοιπόν εφόσον η IP έχει F που δεν είναι καν συνεχής τότε δεν μπορεί να έχει συνάρτηση πυκνότητας.

Π.χ. οι διακριτές κατανομές ή η κατανομή του παραδείγματος 5' δεν έχουν συνάρτηση πυκνότητας αφού σε αυτές η F είναι ασυνεχής.

Συνεπώς η συνάρτηση πυκνότητας δεν υπάρχει πάντοτε. Υπάρχουν όπως βλέπουμε περιπτώσεις κατανομών που δεν έχουν συνάρτηση πυκνότητας

[redacted]

Υπενθύμιση:

* αν $\mathbb{P}$ κατανομή πιθανότητας με αθροιστική F κ' υπάρχει $f:\mathbb{R}\to\mathbb{R}$ τέτοια ώστε $\forall x\in\mathbb{R}$ $F(x)=\int_{-\infty}^{x} f(z)dz$ τότε η f ονομάζεται συνάρτηση πυκνότητας της $\mathbb{P}$.

ΠΡΟΣΟΧΗ: Τυπικά αυτό και για να ακριβολογούμε τον ορισμό της f και χρειάζεται ένας τύπος ολοκληρώματος που διαφέρει από το ολοκλήρωμα Riemann. Δεν ασχολούμαστε με τέτοιες λεπτομέρειες. Ότι συναντήσουμε θα αφορά σε ολοκληρώματα όπως τα ξέρουμε.

* Ύπαρξη: Δεν έχει κάθε $\mathbb{P}$ συνάρτηση πυκνότητας. Π.χ. οι διακριτές κατανομές δεν έχουν αφού οι αθροιστικές τους δεν είναι καν συνεχείς.

Συνέχεια:

Μοναδικότητα: Είναι δυνατόν να αποδειχθεί ότι όταν υπάρχει η f η F είναι παραγωγίσιμη σε κάθε $x\in\mathbb{R}$ εκτός ενδεχομένων από μικρού πλήθους σημεία. Επίσης στα σημεία που είναι παραγωγίσιμη η F έχουμε ότι η $f=\frac{dF}{dx}$. Στα σημεία μη παραγωγισιμότητας της F η f είναι δυνατόν να παίρνει αυθαίρετες τιμές. Οπότε είναι δυνατόν να μην είναι μοναδική. Το παρακάτω μας λέει και το πως να υπολογίζουμε την f όταν γνωρίζουμε ότι υπάρχει. Μέσω παραγώγισης στα σημεία παραγωγισιμότητας της F, και δίνοντας αυθαίρετα τιμές στα σημεία μη παραγωγισιμότητας. (Στα παραδείγματα που θα δούμε παρακάτω θα μας δίνονται συγκεκριμένες εκδοχές της f στις οποίες θα συμφωνούμε, όροι.)

Περαιτέρω Ιδιότητες: (έστω λοιπόν ότι η f υπάρχει)

α. Είναι δυνατόν να αποδειχθεί ότι η f μπορεί να επιλεγεί έτσι ώστε $f(x) \geq 0 \quad \forall x \in \mathbb{R}$.

β. $\displaystyle\int_{-\infty}^{+\infty} f(z)dz = \lim_{x \to +\infty} \int_{-\infty}^{x} f(z)dz = \lim_{x \to +\infty} F(x) = 1.$

* Είναι δυνατόν να αποδειχθεί ότι έχουμε μια συνάρτηση $f: \mathbb{R} \to \mathbb{R}$ που ικανοποιεί τα α, β, τότε η f είναι συνάρτηση πυκνότητας μοναδικής κατανομής $\mathbb{P}$.

Άσκηση: Έστω η $f(x) = \begin{cases} 0, & x \notin [0,1] \\ ce^{x}, & x \in [0,1] \end{cases}$ όπου c σταθερά. Να προσδιοριστεί, αν υπάρχει, τιμή της c για την οποία η f είναι συνάρτηση πυκνότητας για κάποια $\mathbb{P}$.

Βάσει του (*), προκειμένου η f να είναι συνάρτηση πυκνότητας αρκεί να υπάρχει τιμή για το c ώστε να ικανοποιούνται τα α, β. Για να ικανοποιείται το α, αρκεί $c \geq 0$. Για να ικανοποιείται το β θα πρέπει να υπάρχει c τέτοιο ώστε

$\int_{-\infty}^{+\infty} f(z)dz = 1$. Έχουμε $\int_{-\infty}^{+\infty} f(z)dz = \int_{-\infty}^{0} f(z)dz + \int_{0}^{1} f(z)dz$

$+ \int_{1}^{+\infty} f(z)dz = \int_{-\infty}^{0} 0dz + \int_{0}^{1} ce^z dz + \int_{1}^{+\infty} 0dz = c\int_{0}^{1} e^z dz = c\, e^x \Big|_0^1$

$= c[e^1 - e^0] = c(e-1)$

$\int_{-\infty}^{0} 0dz = 0$, $\int_{1}^{+\infty} 0dz = 0$: ορισμένα ολοκληρώματα

Συνεπώς για να ισχύει η β θα πρέπει $c(e-1) = 1 \iff$

$$c = \frac{1}{e-1} > 0.$$

Για να ισχύουν οι α και β ταυτόχρονα θα πρέπει $c = \frac{1}{e-1}$.

Επομένως η $f(x) = \begin{cases} 0, & x \notin [0,1] \\ \frac{1}{e-1} e^x, & x \in [0,1] \end{cases}$ θα είναι συνάρτηση πυκνότητας κάποιας $\mathbb{P}$. $\square$

γ. Επειδή μέσω της f και ολοκλήρωσης μπορούμε να βρίσκουμε την F και επειδή μέσω της F μπορούμε να βρίσκουμε τις πιθανότητες που αποδίδει η $\mathbb{P}$, έχουμε ότι το να γνωρίζουμε την f ισοδυναμεί με το να γνωρίζουμε την $\mathbb{P}$. (Οπότε και μπορούμε να χρησιμοποιούμε την f για να βρίσκουμε τις πιθανότητες που η $\mathbb{P}$ αποδίδει).

Π.χ. αν $\alpha < \beta$, $\mathbb{P}((\alpha,\beta)) \overset{*}{=} \mathbb{P}((\alpha,\beta]) \overset{*}{=} \mathbb{P}([\alpha,\beta)) \overset{*}{=}$

$\overset{*}{=} \mathbb{P}([\alpha,\beta]) = \dots$

Οι ισότητες * προκύπτουν από το ότι: η f υπάρχει $\Rightarrow$ η F είναι αναγκαστικά συνεχής $\Rightarrow$ η F συνεχής στα α, β.

$$\dots = \underline{F(\beta) - F(\alpha)} = \int_{-\infty}^{\beta} f(z)\,dz - \int_{-\infty}^{\alpha} f(z)\,dz = \int_{\alpha}^{\beta} f(z)\,dz.$$

Π.χ. $\mathbb{P}\big((-\infty, \alpha)\big) = \mathbb{P}\big((-\infty, \alpha]\big) = F(\alpha) = \int_{-\infty}^{\alpha} f(z)\,dz.$

προκύπτει όπως παραπάνω

Π.χ. $\mathbb{P}\big([\alpha, +\infty)\big) = \mathbb{P}\big((\alpha, +\infty)\big) = 1 - F(\alpha) = \underset{\text{από την } \alpha}{\int_{-\infty}^{+\infty} f(z)\,dz} - \underset{\text{από τον ορισμό της } f}{\int_{-\infty}^{\alpha} f(z)\,dz}$

$$= \int_{\alpha}^{+\infty} f(z)\,dz.$$

Δηλαδή σε αυτές περιπτώσεις ολοκληρώνουμε την συνάρτηση πυκνότητας στο διάστημα την πιθανότητα του οποίου μας ενδιαφέρει να υπολογίσουμε.

Π.χ. $\mathbb{P}(\{\alpha\}) = 0$ επειδή η f υπάρχει $\Rightarrow$ F συνεχής $\Rightarrow$ F συνεχής στο α

Τέλος Διάλεξης 19

(δείτε κι την τελευταία σελίδα)

δ. Έστω $(\alpha, \beta) \subseteq supp'$ (δηλ. το διάστημα βρίσκεται εκτός ολοκλήρου εκτός του στηρίγματος της $\mathbb{P}$). Επομένως η F σταθερή στο (α,β) και άρα παραγωγίσιμη και συνεπώς $\frac{dF(x)}{dx} = 0$ $\forall x \in (\alpha, \beta)$ οπότε $f(x) = 0$ $\forall x \in (\alpha, \beta)$. Δηλαδή

Εκτός του supp η f θα είναι σταθερή στο 0.

Παραδείγματα.

5. Ομοιόμορφη κατανομή στο $[\alpha,\beta]$ $(\mathrm{Unif}[\alpha,\beta])$

$$- \ \mathrm{supp} = [\alpha,\beta]$$

$$- \ F(x) = \begin{cases} 0, & x<\alpha \\ \dfrac{x-\alpha}{\beta-\alpha}, & \alpha\le x<\beta \\ 1, & \beta\le x \end{cases}$$

Είναι δυνατόν να αποδειχθεί ότι η f υπάρχει. Μας δίνεται η συνήθης εκδοχή της f, η οποία είναι η εξής:

$$f(x) = \begin{cases} 0, & x\notin[\alpha,\beta] \\ \dfrac{1}{\beta-\alpha}, & x\in[\alpha,\beta] \end{cases}$$

6. Εκθετική κατανομή με παράμετρο $\lambda>0$ $(\mathrm{Exp}(\lambda))$.

$$- \ \mathrm{supp} = [0,+\infty)$$

$$- \ F(x) = \begin{cases} 0, & x<0 \\ 1-\exp(-\lambda x), & x\ge 0 \end{cases}$$

Είναι δυνατόν να αποδειχθεί ότι η f υπάρχει. Η συνήθης εκδοχή της f για την εν λόγω κατανομή είναι η

$$f(x) = \begin{cases} 0, & x<0 \\ \lambda e^{-\lambda x}, & x\ge 0. \end{cases}$$

[redacted]

Συνέχεια Παραδειγμάτων:

Παράδειγμα 7. $N(\mu,v)$, $\mu\in\mathbb{R}$, $v>0$

– $supp = \mathbb{R}$

– $F(x) = \int_{-\infty}^{x} f(z)dz$, $f(z) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(z-\mu)^2}{2v}\right)$

Προφανώς η f υπάρχει και εφ' ορισμού είναι η

$$\frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2}\frac{(z-\mu)^2}{v}\right).$$

Για την τυπική κανονική κατανομή ($\mu=0$, $v=1$ – $N(0,1)$)

η συνάρτηση πυκνότητας παίρνει την μορφή:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

$\varphi(-z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(-z)^2\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) = \varphi(z)$ οπότε η

φ είναι άρτια συνάρτηση.

Έστω $\alpha\in\mathbb{R}$, $\mathbb{P}([\alpha,+\infty)) = \int_{\alpha}^{+\infty} \varphi(z)dz =$

κάνουμε την αντικατάστα-
ση $y=-z \Leftrightarrow z=-y$
$dz=-dy$

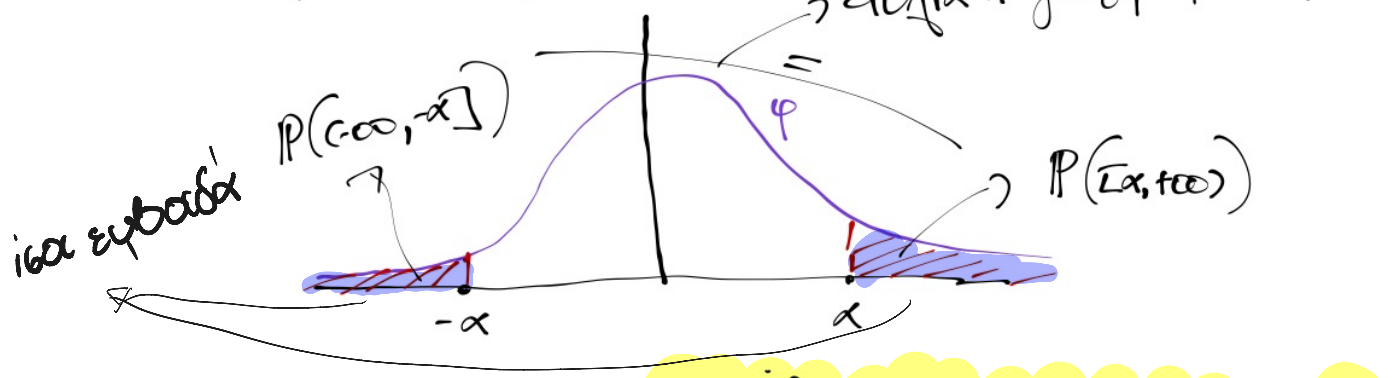
$$= \int_{-\alpha}^{-\infty} \varphi(-y)(-dy) = -\int_{-\alpha}^{-\infty} \varphi(y)dy = \int_{-\infty}^{-\alpha} \varphi(y)dy = \mathbb{P}((-\infty,-\alpha])$$

↳ αρτιότητα της φ

Οπότε $\forall x \in \mathbb{R}$ εξαιτίας της αρτιότητας της φ, για την $N(0,1)$

$$\mathbb{P}([x,+\infty)) = \mathbb{P}((-\infty,-x])$$ (συμμετρία που σχετίζεται με το ότι η φ είναι άρτια).

άθλια η ζωγραφική μου!



Επομένως το παραπάνω είναι παράδειγμα του πώς ιδιότητες της συνάρτησης πυκνότητας (εδώ η αρτιότητα της φ) είναι δυνατόν να αντανακλούν ιδιότητες της κατανομής (εδώ η συμμετρία της $\mathbb{P}$)

Άσκηση. Να δείξετε ότι η ιδιότητα της συμμετρίας ισχύει γενικά για την $N(0,v)$, $v>0$.

Συνέχεια με παράδειγμα της ιδιότητας σ.

π.χ. στην προηγούμενη άσκηση $f(x) = \begin{cases} 0, & x \notin [0,1] \\ \frac{1}{e-1} e^{x}, & x \in [0,1] \end{cases}$

$$\mathbb{P}([0,1/2]) = \int_0^{1/2} f(z)\,dz =$$

$$\int_0^{1/2} \frac{1}{e-1} e^{z}\,dz = \frac{1}{e-1}\int_0^{1/2} e^{z}\,dz = \frac{1}{e-1}\left[e^{z}\Big|_0^{1/2} \right]$$

$$= \frac{1}{e-1}\left(e^{1/2} - e^{0}\right) = \frac{e^{1/2}-1}{e-1}$$

να $\longrightarrow$ $\mathbb{P}((-\infty, 1/3))$

να $\longrightarrow$ $\mathbb{P}([1/2, +\infty))$

να $\longrightarrow$ $\mathbb{P}(\{0\})$

Διάλεξη 19