Διάλεξη 13/05/20

Παράδειγμα 7. Κανονική κατανομή με παραμέτρους $\mu \in \mathbb{R}$ (μέσος) και $v>0$ (διακύμανση) (Normal or Gaussian distribution) – $N(\mu, v)$

– $\text{supp} = \mathbb{R}$ (η πρώτη από αυτές που έχουμε δει που στηρίζεται σε όλο το $\mathbb{R}$)

– $F(x) = \int_{-\infty}^{x} f(z)\,dz$ όπου $f: \mathbb{R} \to \mathbb{R}$ με $(*)$ → καταχρηστικό

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

(η αθροιστική έχει την μορφή ορισμένου ολοκληρώματος ως προς την f)

– Καταρχάς επειδή το ολοκλήρωμα στο $(*)$ είναι καταχρηστικό θα έπρεπε να εξασφαλίσουμε $\forall x \in \mathbb{R}$ $\int_{-\infty}^{x} f(z)\,dz \in \mathbb{R}$ (κ' δεν είναι π.χ. κάποια απροσδιοριστία ή κάποιο άπειρο). Έτσι θα εξασφαλίσουμε ότι η F παραπάνω είναι συνάρτηση $\mathbb{R} \to \mathbb{R}$.

– Παρατηρούμε ότι $f(z)>0$, επομένως το $\int_{-\infty}^{x} f(z)\,dz$ θα υπάρχει ως πραγματικός αριθμός αν υπάρχει το $\int_{-\infty}^{+\infty} f(z)\,dz$.

Έχουμε

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

$\frac{1}{\sqrt{\pi}\sqrt{2v}}\,dz \to dy$ $\qquad$ $\left(\frac{z-\mu}{\sqrt{2v}}\right)^2 \to y^2$

Μας δίνεται το ολοκλήρωμα Gauss $\int_{-\infty}^{+\infty} \exp(-y^2)\,dy = \sqrt{\pi}$

Θέτουμε $y = \frac{z-\mu}{\sqrt{2\nu}}$, $dy = \frac{1}{\sqrt{2\nu}}dz$, επομένως στο (**) έχουμε

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \exp(-y^2)\,dy = \frac{1}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{+\infty} \exp(-y^2)\,dy}_{\sqrt{\pi}} = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1$$

Όταν $z \to -\infty$, $y = \frac{z-\mu}{\sqrt{2\nu}} \to -\infty$ $\quad \sqrt{2\nu} > 0$

$z \to +\infty$, $y = \frac{z-\mu}{\sqrt{2\nu}} \to +\infty$

Επομένως $\int_{-\infty}^{+\infty} f(z)\,dz \in \mathbb{R} \Rightarrow \int_{-\infty}^{x} f(z)\,dz \in \mathbb{R}$ $\forall x \in \mathbb{R}$ (επειδή $f > 0$)

Επομένως η F είναι πραγματική συνάρτηση $\mathbb{R} \to \mathbb{R}$.

Είναι όμως αθροιστική συνάρτηση;

Παρατηρούμε: * η F παραγωγίσιμη $\forall x \in \mathbb{R}$ αφού

$$\frac{dF(x)}{dx} = \frac{d\int_{-\infty}^{x} f(z)\,dz}{dx} = f(x) = \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(x-\mu)^2}{2\nu}}$$

Οπότε είναι συνεχής.

* η F είναι γνησίως αύξουσα αφού

$$\frac{dF(x)}{dx} = f(x) > 0 \quad \forall x \in \mathbb{R}$$

Επομένως η F είναι συνεχής (επομένως κ' από δεξιά συνεχής) κ' είναι κ' γνησίως αύξουσα (άρα αύξουσα)

Συνεπώς οι ιδιότητες i. και iii. ικανοποιούνται.

Τι συμβαίνει με την ii;

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

από τον παραπάνω υπολογισμό.

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty}\int_{-\infty}^{x} f(z)\,dz = \dots = 0.$$

Είναι δυνατόν να αποδειχθεί (αν θέλετε κάντε το ως άσκηση)

Επομένως ικανοποιείται και η ii. Άρα πράγματι η F είναι καλώς ορισμένη αθροιστική συνάρτηση που βάσει του θεωρήματος χαρακτηρισμού θα αντιστοιχεί μοναδική $\mathbb{P}$ στο $\mathbb{R}$, την $N(\mu,\nu)$.

- Είναι δυνατόν να αποδειχθεί ότι η F μπορεί να εκφρασθεί μόνο με μορφή ολοκληρώματος.
- Η $N(\mu,\nu)$ είναι συνεχής αφού το στήριγμά της είναι διάστημα ($supp = \mathbb{R}$)
- Η αθροιστική της είναι συνεχής οπότε η $\mathbb{P}(\{x\})=0 \ \forall x\in\mathbb{R}$.
- $\underbrace{\mathbb{P}([\alpha,\beta]) = \mathbb{P}((\alpha,\beta)) = \mathbb{P}([\alpha,\beta)) = \mathbb{P}((\alpha,\beta])}_{\text{συνέχεια της } F} =$

$$= F(b) - F(a) = \int_{-\infty}^{b} f(z)\,dz - \int_{-\infty}^{a} f(z)\,dz = \int_{a}^{b} f(z)\,dz =$$

$$= \frac{1}{\sqrt{2\pi v}} \int_{a}^{b} \exp\left(-\frac{(z-\mu)^2}{2v}\right) dz \quad \text{επομένως}$$

για την $N(\mu,v)$ οι πιθανότητες που αποδίδονται σε διαστήματα πεπερασμένης ακτίνας προκύπτουν ως ολοκληρώματα σε αυτά της f.

— Επειδή η F μέσω της f εξαρτάται μονοσήμαντα από το $\binom{\mu}{v}$ σε κάθε τιμή του $\binom{\mu}{v}$ αντιστοιχεί κ μία και μοναδική $N(\mu,v)$. Συνεπώς στα παραπάνω ορίσαμε την οικογένεια των κανονικών κατανομών. Ιδιαίτερο παράδειγμα αυτής της οικογένειας είναι η τυπική κανονική κατανομή (Standard Normal) που προκύπτει για $\mu=0$ κ' $v=1$ $N(0,1)$. Συνήθως για αυτή έχουμε τους συμβολισμούς:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(z)\,dz, \quad \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$\square$

Σε αυτή την παράγραφο είδαμε ότι η αθροιστική συνάρτηση υπάρχει και είναι μοναδική για κάθε κατανομή πιθανότητας στο $\mathbb{R}$ και την αναπαριστά.

Β. 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 αντιληφθούμε 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 $\mathbb{P}$ 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