Παραδείγματα

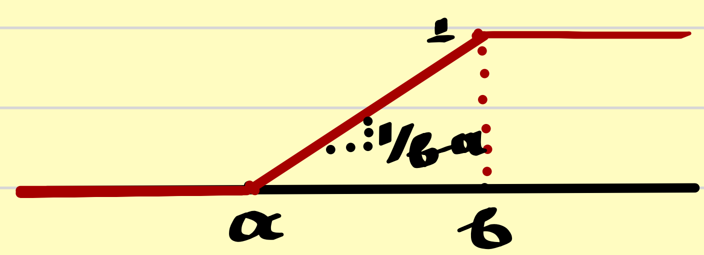
Εφόσον όπως είδαμε η cdf αναπαριστά την P, είναι δυνατόν να ορίζουμε την P δίνοντας απλώς την F. Σε κάθε ένα από τα παρακάτω θα δίνεται η F (και χωρίς να χρειάζεται το supp - γιατί δεν είναι αναγκαίο να δίνεται το supp;) και θα ελέγχεται το κατά πόσο η F ικανοποιεί τις τρεις ιδιότητες του θεωρήματος. Εφόσον τις ικανοποιεί τότε θα ορίζει μονοσήμαντα κατανομή πιθανότητας.

1. Ομοιόμορφη (Uniform) κατανομή στο $[\alpha,\beta]$, όπου $\alpha,\beta\in\mathbb{R}$, με $\alpha<\beta$ ($Unif_{[\alpha,\beta]}$)

$$supp=[\alpha,\beta]$$

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

Το γράφημα της F θα μοιάζει με το παρακάτω:



Είναι προφανές ότι η παραπάνω ικανοποιεί τις τρεις ιδιότητες (είναι μάλιστα συνεχής - γιατί; Δείξτε το!) και συνεπώς ορίζει κατανομή που ονομάζεται ομοιόμορφη στο $[\alpha,\beta]$.
Επεκτείνοντας το παραπάνω ορίζει την οικογένεια των ομοιόμορφων κατανομών, αφού έχουμε μια διαφορετική ομοιόμορφη κατανομή, για κάθε διαφορετική τιμή (α,β) με $\alpha<\beta$. Συνεπώς η οικογένεια αυτή έχει το ίδιο πλήθος στοιχείων με το πλήθος των δυνατών διανυσμάτων (α,β) όπου η πρώ-

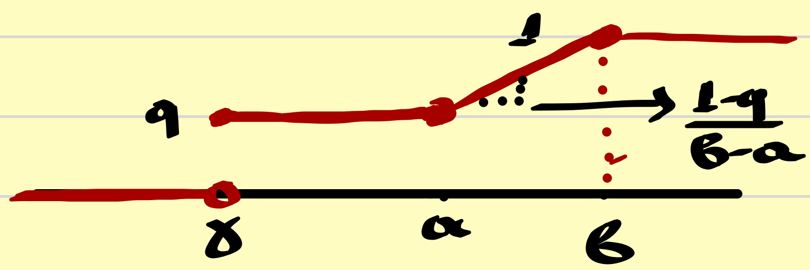
τη συνιστώσα είναι μικρότερη της δεύτερης (γιατί;).*
Σημειώνουμε τέλος ότι κάθε μία ομοιόμορφη είναι παράδειγμα συνεχούς κατανομής (γιατί;) με συνεχές cdf. Επομένως, $\forall x\in[\alpha,\beta]$, $P(\{x\})=0$ (γιατί;). □

2. Έστω $q\in(0,1)$, $\alpha,\beta,\gamma\in\mathbb{R}$ με $\gamma<\alpha<\beta$. Έστω η κατανομή P που ορίζεται από:

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

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

Το γράφημα της F μοιάζει με το παρακάτω



Είναι προφανές από τον ορισμό (και το γράφημα) ότι η F ικανοποιεί και τις τρεις ιδιότητες (δείξτε το!) συνεπώς ορίζει μονοσήμαντα κατανομή P στο $\mathbb{R}$. Παρατηρούμε ότι αποτελεί παράδειγμα μεικτής κατανομής αφού $\mathrm{supp}=\{\gamma\}\cup[\alpha,\beta]$ και $\gamma<\alpha$, ενώ $P(\{\gamma\})=F(\gamma)-\lim_{x\to\gamma^-}F(x)=q-0=q>0$.

$\forall x\in\mathbb{R}$, $x\neq\gamma$ έχουμε ότι $P(\{x\})=0$ (γιατί;). Παρατηρούμε ότι όταν $q\to 0$ έχουμε ότι η F συγκλίνει στο cdf της $Unif_{[\alpha,\beta]}$ $\forall x\in\mathbb{R}$, ενώ όταν $q\to 1$ η F συγκλίνει στο cdf της εκφυλισμένης κατανομής στο γ, $\forall x\in\mathbb{R}$ (γιατί;). Επομένως οι δύο προαναφερθείσες, αποτελούν "οριακές" περιπτώσεις της μετατόπισης. □

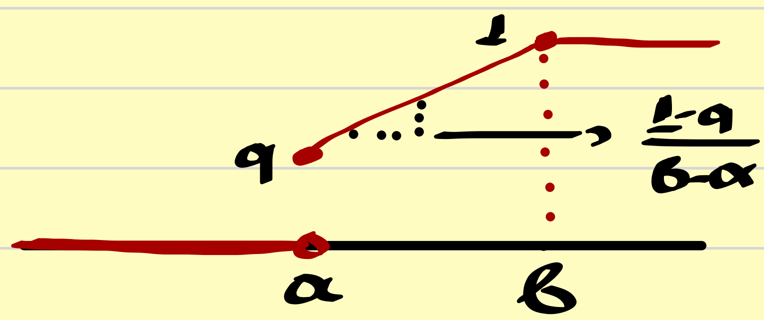
3. $q\in(0,1)$, $\alpha,\beta\in\mathbb{R}$ με $\alpha<\beta$. Έστω η κατανομή P που ορίζεται από:

* Η οικογένεια είναι το σύνολο $\{Unif_{[\alpha,\beta]},\ \alpha,\beta\in\mathbb{R},\ \alpha<\beta\}$.

$$\text{supp} = [\alpha, \beta],$$

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

Το γράφημα της F μοιάζει με αυτό του παρακάτω σχήματος



Και τώρα είναι εύκολο να δειχθεί ότι η F ικανοποιεί και τις τρεις ιδιότητες (δείξτε το!), συνεπώς ορίζει μονοδιάστατη κατανομή P στο $\mathbb{R}$. Αυτή είναι συνεχής αφού $\text{supp}=[\alpha,\beta]$, αλλά έχει ασυνεχή (εξαιτίας του ότι $q \neq 0$) cdf. Συνεπώς δεν πρέπει να συγχέεται το ζήτημα της συνέχειας της P (που αφορά στην μορφή του supp), με το ζήτημα της συνέχειας της F. Έχουμε ότι $P(\{\alpha\}) = F(\alpha) - \lim_{x \to \alpha^-} F(x) =$

$= q + \frac{1-q}{\beta-\alpha}(\alpha-\alpha) - 0 = q$, ενώ $P(\{x\}) = 0, \forall x \in \mathbb{R}, x \neq \alpha$. Παρατηρούμε ότι όταν $q \to 0$ η F συγκλίνει στο cdf της $\text{Unif}[\alpha,\beta]$ $\forall x \in \mathbb{R}$, ενώ όταν $q \to 1$ η F συγκλίνει στο cdf της εκφυλισμένης κατανομής στο α, $\forall x \in \mathbb{R}$. Επομένως οι δύο προαναφερθείσες αποτελούν ειδικές περιπτώσεις της μελετούμενης. □

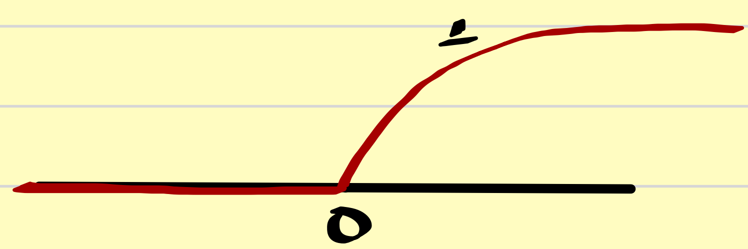
Στα προηγούμενα το supp ήταν φραγμένο. Στα επόμενα αυτό αναιρείται.

4. Έστω $\lambda > 0$. Η εκθετική κατανομή που αντιστοιχεί στην παράμετρο λ, $\text{Exp}(\lambda)$, ορίζεται από:

$$\text{supp} = \mathbb{R}_+ := [0, +\infty)$$

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

Μέσω αυτής μελέτης της F είναι δυνατόν να δούμε ότι το γράφημα της μοιάζει με το γράφημα στο παρακάτω σχήμα:



Έχουμε ότι $\lim_{x\to+\infty}[1-\exp(-\lambda x)] = 1-\lim_{x\to+\infty}\exp(-\lambda x) = 1$, ενώ επειδή εξ υποθέσεως $\lambda>0$, ενώ η $x\to e^x$ είναι αύξουσα, και $e^{-\lambda x}<1, \forall x>0$, $F(x_2)-F(x_1) = \begin{cases} 0, & x_1<x_2\leq 0 \\ 1-e^{-\lambda x_2}, & x_1\leq 0<x_2 \\ e^{-\lambda x_1}-e^{-\lambda x_2}, & 0<x_1<x_2 \end{cases} \geq 0$. Συνεπώς (γιατί; Συμπληρώστε!) οι τρεις ιδιότητες ικανοποιούνται και άρα η εν λόγω P είναι καλώς ορισμένη. Και εδώ έχουμε παράδειγμα συνεχούς P με συνεχή cdf. Επομένως, $\forall x\in\mathbb{R}$, $P(\{x\})=0$. Αν σκεφτούμε τα παραπάνω και ως συνάρτηση του λ τότε αποκτούμε την οικογένεια των εκθετικών κατανομών $\{\mathrm{Exp}(\lambda), \lambda>0\}$ το πλήθος των μελών της οποίας ταυτίζεται (γιατί;) με το πλήθος των στοιχείων του $(0,+\infty)$. Έτσι π.χ. αν $0<\lambda_1<\lambda_2$, για την $\mathrm{Exp}(\lambda_1)$ έχουμε ότι $P([0,1]) = 1-\exp(-\lambda_1)-(1-\exp(-\lambda_1 0)) = 1-\exp(-\lambda_1)$ ενώ αναλόγως για την $\mathrm{Exp}(\lambda_2)$ έχουμε αναλόγως $P([0,1]) = 1-\exp(-\lambda_2)$. $\lambda_1<\lambda_2 \Rightarrow \exp(-\lambda_1)>\exp(-\lambda_2) \Rightarrow 1-\exp(-\lambda_1)<1-\exp(-\lambda_2)$ και επομένως η $\mathrm{Exp}(\lambda_1)$ αποδίδει μικρότερη πιθανότητα στο $[0,1]$ από αυτήν που αποδίδει η $\mathrm{Exp}(\lambda_2)$. (Προσπαθήστε να δείξετε το παραπάνω χρησιμοποιώντας την παραγωγισιμότητα της F ως προς λ, όταν $x\geq 0$, αποκτώντας έτσι παράδειγμα του πως "αναλυτικές" ιδιότητες της F

ως μέρος της παραμέτρου που ορίζει την οικογένεια κατανομών ιδιότητες των μελών της). $\square$

Στο τελευταίο παράδειγμα το supp είναι το μεγαλύτερο δυνατό.

5. Έστω $\mu \in \mathbb{R}$, $v>0$. Η κανονική κατανομή με μέσο μ, και διακύμανση v $(N(\mu, v))$, ορίζεται από:

$$\text{supp} = \mathbb{R}$$

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

$$= \frac{1}{\sqrt{2\pi v}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2v}(x-\mu)^2\right) dz.$$

Παρατηρούμε σε αυτό το παράδειγμα έχουμε ότι

$$\forall x \in \mathbb{R},\ F(x) = \int_{-\infty}^{x} f(z)\,dz \quad \text{για } f:\mathbb{R}\to\mathbb{R} \quad \text{με}$$

$f(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}(x-\mu)^2\right)$. Με το γιατί αυτό μπορεί να είναι δυνατό θα ασχοληθούμε αργότερα (Πρβλ. την έννοια της συνάρτησης πυκνότητας). Αν ισχύει ότι (*) $\int_{-\infty}^{x} f(z)dz \in \mathbb{R}, \forall x \in \mathbb{R}$ τότε εξουσίας σχετικών ιδιοτήτων μπορεί να αποδειχθεί ότι $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \int_{-\infty}^{x} f(z)dz = 0$, όπως επίσης και ότι η F συνεχής σε κάθε $x \in \mathbb{R}$.

Επίσης, αν ισχύει η (*) και $x_1 < x_2$, έχουμε ότι $F(x_2) - F(x_1) = \left(\int_{-\infty}^{x_2} - \int_{-\infty}^{x_1}\right)(f(z)dz) = \int_{x_1}^{x_2} f(z)dz > 0$ αφού $f(z) > 0\ \forall z \in \mathbb{R}$

και $x_1<x_2$. Επειδή $f(z)>0\ \forall z\in\mathbb{R}$ είναι δυνατόν να δειχθεί ότι η (*) θα ισχύει αν $\int_{-\infty}^{+\infty} f(z)dz\in\mathbb{R}$. Επίσης αν $\int_{-\infty}^{+\infty} f(z)dz=1$ τότε και $\lim_{x\to+\infty} F(x)=1$. Επομένως αν δείξουμε το τελευταίο, τότε θα έχουμε ότι ισχύουν και οι τρεις ιδιότητες και συνεπώς η $N(\mu,v)$ θα είναι καλώς ορισμένη από την παραπάνω F.

Μας δίνεται το ολοκλήρωμα του Gauss ως: $\int_{-\infty}^{+\infty}\exp(-\rho^2)d\rho=\sqrt{\pi}$

(είναι εκτός του εύρους του μαθήματος, αλλά ένας τρόπος να αποδείξουμε το παραπάνω είναι μέσω της χρήσης πολικών συντεταγμένων).

Έχουμε ότι
$$\int_{-\infty}^{+\infty} f(z)dz=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2v}(z-\mu)^2\right)dz$$
$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi v}}\exp\left(-\left(\frac{z-\mu}{\sqrt{2v}}\right)^2\right)dz=\left(\rho=\frac{z-\mu}{\sqrt{2v}},\ d\rho=\frac{1}{\sqrt{2v}}dz\right)$$
$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}\exp(-\rho^2)d\rho=\frac{\sqrt{\pi}}{\sqrt{\pi}}=1.$$
[Εξαιτίας του παραπάνω έχουμε ότι η $N(\mu,v)$ ονομάζεται και Gaussian κατανομή]. Επίσης αν αντιληφθούμε την F και ως συνάρτηση του (μ,v), τότε αποκτάμε την οικογένεια των κανονικών κατανομών $\{N(\mu,v),\ \mu\in\mathbb{R}, v>0\}$ που έχει πλήθος στοιχείων, όσο και το πλήθος των διανυσμάτων της μορφής (μ,v) με $\mu\in\mathbb{R}, v>0$. □

Παρατηρήσεις.

1. Τα παραπάνω κάνουν εμφανές το γιατί το supp είναι εξ ορισμού κλειστό. Εάν δεν ήταν κλειστό τότε γενικά η μοναδικότητά του θα χάνονταν (γιατί;).
2. Στο παράδειγμα 1 η $Unif_{[0,1]}$ ονομάζεται τυπική ομοιόμορφη (standard Uniform).

3. Αναλόγως, στο παράδειγμα 5, η $N(0,1)$ ονομάζεται τυπική κανονική (standard Normal). Στην ειδική αυτή περίπτωση η cdf συμβολίζεται με Φ και έχουμε ότι:

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

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.