Curs 5

Cristian Niculescu

1 Galerie a variabilelor aleatoare continue

1.1 Scopurile învăţării

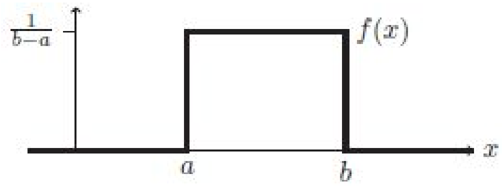
1. Să poată da exemple de ceea ce modelează repartiţiile uniformă, exponenţială şi normală.
2. Să poată da domeniile de valori şi pdf-urile repartiţiilor uniformă, exponenţială şi normală.

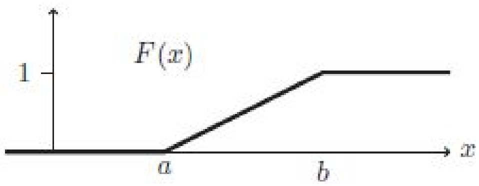
1.2 Introducere

Introducem câteva repartiţii continue funadmentale. Pentru fiecare repartiţie dăm domeniul de valori, pdf, cdf şi o scurtă descriere a situaţiilor pe care le modelează. Toate aceste repartiţii depind de parametri, pe care îi specificăm.
Cu toate că o abordăm spre final, repartiţia normală este cea mai importantă din cele definite aici. Când dăm funcţia de repartiţie (cdf, prescurtat "repartiţia") se subînţelege, dacă este cazul, că la stânga intervalului pe care este dată este 0, iar la dreapta este 1.

1.3 Repartiţia uniformă

1. Parametri: $a < b$.
2. Domeniu de valori: $[a, b]$.
3. Notaţie: uniform(a, b) sau $U(a, b)$.
4. Densitate: $f(x) = \frac{1}{b-a}$ pentru $a \leq x \leq b$.
5. Repartiţia: $F(x) = (x - a)/(b - a)$ pentru $a \leq x \leq b$.
6. Modele: Toate rezultatele din domeniul de valori au probabilitate egală (mai precis, toate rezultatele au aceeaşi densitate de probabilitate).
Grafice pentru pdf şi cdf:





Exemple. 1. Presupunem că avem o ruletă marcată în milimetri. Dacă măsurăm (până la cel mai apropiat marcaj) lungimea unor articole care au aproximativ 1 metru, eroarea de rotunjire va fi repartizată uniform între -0.5 şi 0.5 milimetri.
2. Multe jocuri de tablă folosesc săgeţi care se învârt pentru a introduce hazardul. Când este învârtită, săgeata se opreşte la un unghi care este repartizat uniform între 0 şi 2π radiani.
3. În cele mai multe generatoare de numere pseudo-aleatoare, generatorul de bază simulează o repartiţie uniformă şi toate celelalte repartiţii sunt construite transformând generatorul de bază.

1.4 Repartiţia exponenţială

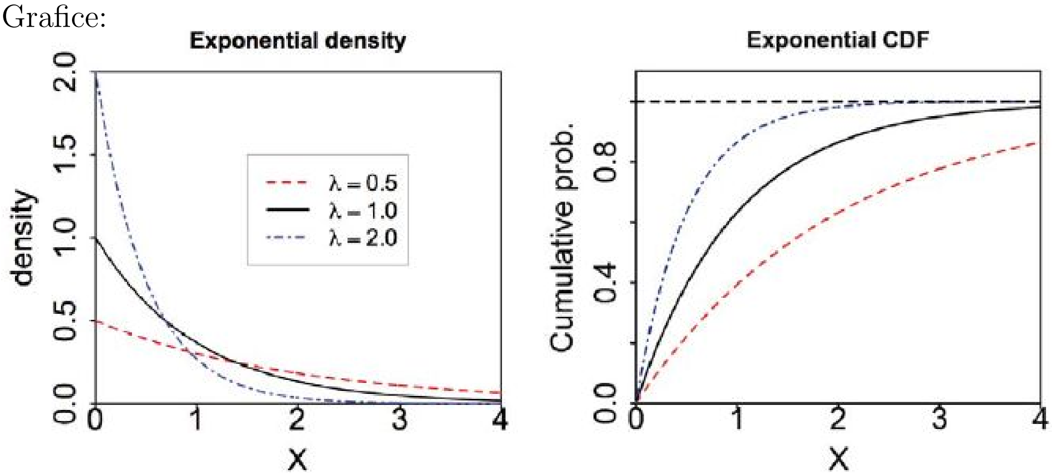
1. Parametru: $\lambda > 0$.
2. Domeniu de valori: $[0, \infty)$.
3. Notaţie: exponential(λ) sau $exp(\lambda)$.
4. Densitate: $f(x) = \lambda e^{-\lambda x}$ pentru $x \geq 0$.
5. Repartiţie: $F(x) = 1 - e^{-\lambda x}$ pentru $x \geq 0$.
6. *Repartiţia cozii drepte*: $P(X > x) = 1 - F(x) = e^{-\lambda x}$.
7. Modele: Timpul de aşteptare pentru un proces continuu de schimbare a stării.

Exemple. 1. Dacă ies afară după curs şi aştept un taxi, timpul meu de aşteptare în minute este repartizat exponenţial. În acest caz λ este dat de 1 supra numărul mediu de taxiuri care trec pe minut (în această perioadă de timp din zilele lucrătoare).
2. Repartiţia exponenţială modelează timpul de aşteptare până când un izotop instabil suferă o descompunere nucleară. În acest caz λ este legat de timpul de înjumătăţire al izotopului.
Proprietatea de lipsă a memoriei: Sunt şi alte repartiţii care modelează timpii de aşteptare, dar repartiţia exponenţială are proprietatea adiţională că este lipsită de memorie. Iată ce înseamnă aceasta în contextul exemplului 1. Presupunem că probabilitatea ca un taxi să sosească în primele 5 minute este p. Dacă aştept 5 minute şi de fapt nu soseşte niciun taxi, atunci probabilitatea ca un taxi să sosească în următoarele 5 minute este tot p.
Din contra, să presupunem că merg la o staţie de metrou şi aştept următorul tren. Deoarece trenurile sunt coordonate să urmeze un program (de exem-

plu, cel mult 12 minute între trenuri), dacă aştept 5 minute fără să văd un tren, atunci este o probabilitate mult mai mare ca trenul să sosească în următoarele 5 minute. În particular, timpul de aşteptare pentru metrou nu este fără memorie şi un model mai bun pentru el ar fi repartiţia uniformă pe intervalul [0,12].
Lipsa de memorie a repartiţiei exponenţiale este analoagă lipsei de memorie a repartiţiei geometrice (discrete), unde faptul că am aruncat 5 reversuri la rând nu ne dă nicio informaţie despre următoarele 5 aruncări.
Într-adevăr, repartiţia exponenţială este precis perechea continuă a repartiţiei geometrice, care modelează timpul de aşteptare pentru un proces discret de schimbare de stări. Mai formal, lipsa de memorie înseamnă că probabilitatea de a aştepta mai mult de t minute este neafectată de faptul că am aşteptat deja s minute fără ca evenimentul să se producă. În simboluri, $P(X > s + t|X > s) = P(X > t)$.
Demonstraţia lipsei de memorie: Deoarece $(X > s + t) \cap (X > s) = (X > s + t)$, avem

$$P(X > s + t|X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t), \text{q.e.d.}$$

Grafice:



1.5 Repartiţia normală

În 1809, Carl Friedrich Gauss a publicat o monografie introducând câteva noţiuni care au devenit fundamentale în statistică: repartiţia normală, estimarea de verosimilitate maximă şi metoda celor mai mici pătrate. Din acest motiv, repartiţia normală mai este numită şi repartiţia *Gaussiană.* Ea este cea mai importantă repartiţie continuă.
1. Parametri: $\mu \in \mathbb{R}, \sigma > 0$.

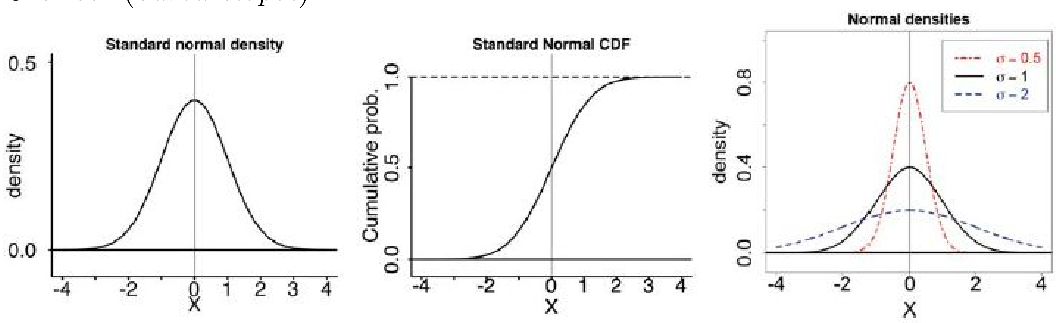
2. Domeniul de valori: $\mathbb{R}$.
3. Notaţie: normal(μ, σ^2) sau $N(\mu, \sigma^2)$.
4. Densitate: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$.
5. Repartiţie: $F(x)$ nu are formulă, aşa că folosim tabele sau comenzi ca `pnorm` în R pentru a calcula $F(x)$.
6. Modele: măsurarea erorii, inteligenţei/abilităţii, înălţimii, mediilor loturilor de date.

Repartiţia normală standard $N(0, 1)$ are media 0 şi dispersia 1. Rezervăm Z pentru o variabilă aleatoare normală standard, $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ pentru densitatea normală standard şi $\Phi(z)$ pentru repartiţia normală standard.

Observaţie: Media şi dispersia variabilelor aleatoare continue au aceeaşi interpretare ca în cazul discret. Repartiţia normală $N(\mu, \sigma^2)$ are media μ, dispersia σ^2 şi deviaţia standard σ.

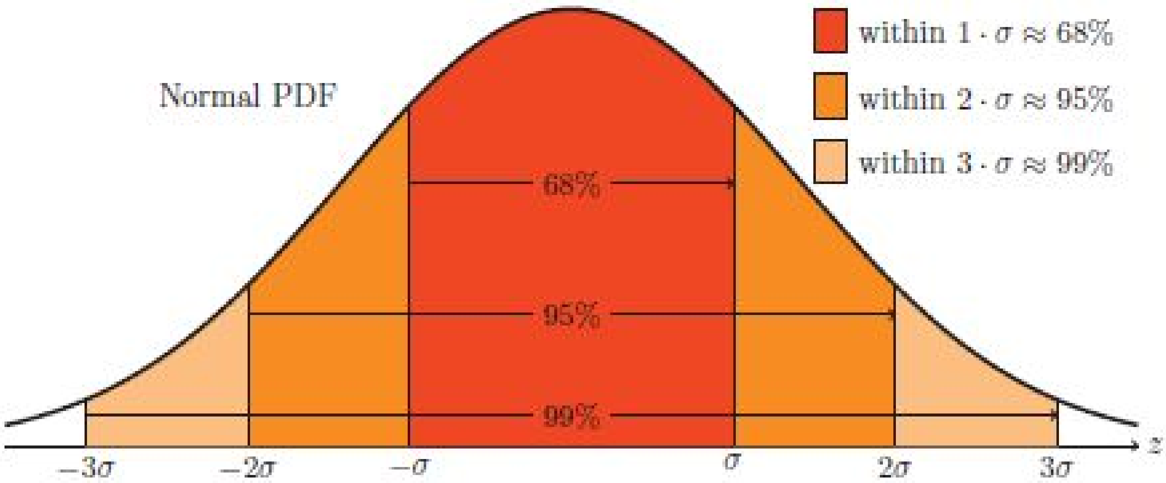
Iată câteva grafice ale repartiţiei normale. Observăm că densităţile au graficele curbe în formă de *clopot*. Mai observăm că pe măsură ce σ creşte, ele devin din ce în ce mai împrăştiate.

Grafice: (*curba clopot*):

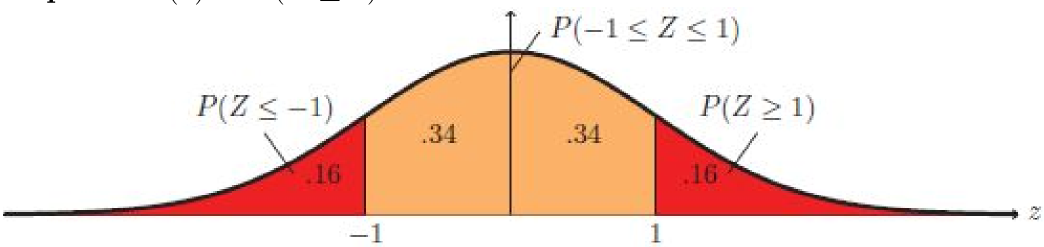


1.5.1 Probabilităţi normale

Pentru a face aproximări este util să ne reamintim următoarea regulă a degetului mare pentru 3 probabilităţi aproximative
$P(-1 \le Z \le 1) \approx 0.68$, $P(-2 \le Z \le 2) \approx 0.95$, $P(-3 \le Z \le 3) \approx 0.99$.



Calcule cu simetrie
Putem folosi simetria graficului densităţii normale standard faţă de dreapta $x = 0$ pentru a face unele calcule.
Exemplul 1. Regula degetului mare spune $P(-1 \leq Z \leq 1) \approx 0.68$. Folosiţi aceasta pentru a estima $\Phi(1)$.
Răspuns: $\Phi(1) = P(Z \leq 1)$.



În figură, cele 2 cozi (în roşu) au împreună aria $1 - 0.68 = 0.32$. Din simetrie, coada stângă are aria 0.16 (jumătate din 0.32), deci $P(Z \leq 1) \approx 0.16 + 0.68 = 0.84$.

1.5.2 Folosirea lui R pentru a calcula $\Phi(z)$

Folosim funcţia R pnorm(x, μ, σ) pentru a calcula $F(x)$ pentru $N(\mu, \sigma^2)$.

```
pnorm(1,0,1)
[1] 0.8413447
pnorm(0,0,1)
[1] 0.5
pnorm(1,0,2)
[1] 0.6914625
pnorm(1,0,1)-pnorm(-1,0,1)
[1] 0.6826895
```

```
pnorm(5,0,5)-pnorm(-5,0,5)
[1] 0.6826895
# Desigur, z poate fi un vector de valori:
pnorm(c(-3,-2,-1,0,1,2,3),0,1)
[1] 0.001349898 0.022750132 0.158655254 0.500000000 0.841344746 0.977249868
[7] 0.998650102
```

Observaţie: Funcţia R pnorm(x, μ, σ) utilizează σ în timp ce notaţia noastră pentru repartiţia normală $N(\mu, \sigma^2)$ foloseşte σ^2.

Iată un tabel de valori cu o acurateţe cu mai puţine zecimale:

z:	-2	-1	0	.3	.5	1	2	3
$\Phi(z)$:	0.0228	0.1587	0.5000	0.6179	0.6915	0.8413	0.9772	0.9987

Exemplul 2. Utilizaţi R pentru a calcula $P(-1.5 \leq Z \leq 2)$.
Răspuns: $P(-1.5 \leq Z \leq 2) = \Phi(2) - \Phi(-1.5) =$ `pnorm(2,0,1)-pnorm(-1.5,0,1)=0.9104427`.

1.6 Repartiţia Pareto

1. Parametri: $m > 0$ şi $\alpha > 0$.
2. Domeniul de valori: $[m, \infty)$.
3. Notaţie: Pareto(m, α).
4. Densitate: $f(x) = \frac{\alpha m^\alpha}{x^{\alpha+1}}$.
5. Repartiţia: $F(x) = 1 - \frac{m^\alpha}{x^\alpha}$, pentru $x \geq m$.
6. Repartiţia coadă: $P(X > x) = m^\alpha / x^\alpha$, pentru $x \geq m$.
7. Modele: Repartiţia Pareto modelează o **lege de putere**, unde probabilitatea ca un eveniment să aibă loc variază ca o putere a unui atribut al evenimentului. Multe fenomene urmează o lege de putere, ca mărimea meteoriţilor, nivelurile veniturilor într-o populaţie şi nivelurile populaţiei în oraşe. Vezi `http://en.wikipedia.org/wiki/Pareto_distribution#Applications`.

2 Manipularea variabilelor aleatoare continue

2.1 Scopul învăţării

Să poată să afle pdf şi cdf ale unei variabile aleatoare definite în termeni de o variabilă aleatoare cu pdf şi cdf cunoscute.

2.2 Transformări de variabile aleatoare

Dacă $Y = aX + b$, atunci proprietăţile mediei şi dispersiei ne spun că $E(Y) = aE(X) + b$ şi $Var(Y) = a^2 Var(X)$. Dar care este funcţia de

repartiţie a lui Y? Dacă Y este continuă, care este pdf a ei?
Adesea, pentru transformarea variabilelor aleatoare discrete, lucrăm cu tabele. Pentru variabile aleatoare continue transformarea pdf este chiar schimbarea de variabilă de la analiza matematică. Transformarea cdf foloseşte direct definiţia cdf.
Reamintim:
1. Cdf a lui X este $F_X(x) = P(X \leq x)$.
2. Pdf a lui X este legată de F_X prin $f_X(x) = F'_X(x)$.
Exemplul 1. Fie $X \sim U(0,2)$, deci $f_X(x) = 1/2$ şi $F_X(x) = x/2$ pe [0,2]. Care sunt domeniul de valori, pdf şi cdf ale lui $Y = X^2$?
Răspuns: Domeniul de valori al lui Y este [0,4].
Pentru a afla cdf folosim definiţia:

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = F_X(\sqrt{y}) = \sqrt{y}/2.$$

Pentru a afla pdf derivăm cdf:

$$f_Y(y) = F'_Y(y) = \left(\frac{\sqrt{y}}{2}\right)' = \frac{1}{4\sqrt{y}}.$$

Un mod alternativ de a afla pdf este prin schimbare de variabilă. Trucul aici este să ne reamintim că $f_X(x)dx$ dă probabilitatea ($f_X(x)$ este densitatea de probabilitate). Iată calculul:

$$y = x^2 \Rightarrow dy = 2xdx \Rightarrow dx = \frac{dy}{2\sqrt{y}}$$

$$f_X(x)dx = \frac{dx}{2} = \frac{dy}{4\sqrt{y}} = f_Y(y)dy.$$

De aceea, $f_Y(y) = \frac{1}{4\sqrt{y}}$.
Exemplul 2. Fie $X \sim exp(\lambda)$, deci $f_X(x) = \lambda e^{-\lambda x}$ pe $[0, \infty)$. Care este densitatea lui $Y = X^2$?
Răspuns: Folosim schimbarea de variabilă:

$$y = x^2 \Rightarrow dy = 2xdx \Rightarrow dx = \frac{dy}{2\sqrt{y}}$$

$$f_X(x)dx = \lambda e^{-\lambda x}dx = \lambda e^{-\lambda\sqrt{y}}\frac{dy}{2\sqrt{y}} = f_Y(y)dy.$$

De aceea, $f_Y(y) = \frac{\lambda}{2\sqrt{y}}e^{-\lambda\sqrt{y}}$.
Exemplul 3. Presupunem $X \sim N(5, 3^2)$. Arătaţi că $Z = \frac{X-5}{3}$ este normală standard, i.e., $Z \sim N(0,1)$.

Răspuns: Din nou folosind schimbarea de variabilă şi formula lui $f_X(x)$ avem:

$$z = \frac{x-5}{3} \Rightarrow dz = \frac{dx}{3} \Rightarrow dx = 3dz$$
$$f_X(x)dx = \frac{1}{3\sqrt{2\pi}} e^{-(x-5)^2/(2\cdot 3^2)} dx = \frac{1}{3\sqrt{2\pi}} e^{-z^2/2} 3dz = f_Z(z)dz.$$

De aceea, $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. Deoarece aceasta este exact densitatea pentru $N(0,1)$, am arătat că Z este normală standard.
Acest exemplu a arătat o proprietate generală importantă a variabilelor aleatoare normale pe care o dăm în exemplul următor.
Exemplul 4. Presupunem $X \sim N(\mu, \sigma^2)$. Arătaţi că $Z = \frac{X-\mu}{\sigma}$ este normală standard, i.e., $Z \sim N(0,1)$.
Răspuns. Este exact acelaşi calcul ca la exemplul trecut cu μ înlocuind pe 5 şi σ pe 3:

$$z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{dx}{\sigma} \Rightarrow dx = \sigma dz$$
$$f_X(x)dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\cdot\sigma^2)} dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2} \sigma dz = f_Z(z)dz.$$

De aceea, $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. Aceasta arată că Z este normală standard.