Matematikai és számítástudományi ismeretek, 1. tétel

1 Első rész

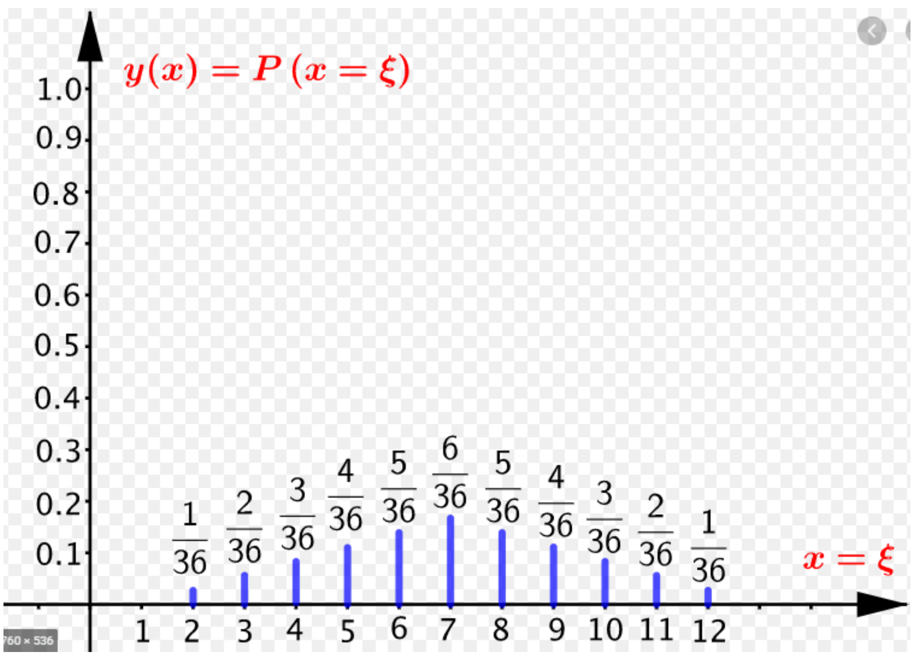
Diszkrét és folytonos valószínűségi eloszlás fogalma. Nevezetes eloszlások: binomiális, Poisson, egyenletes, exponenciális, normális.

Egy $p1, p2, ...$ sorozatot **valószínűségi eloszlás**nak nevezünk, ha $p_i \geq 0$, és $\sum_{i=1}^{\infty} p_i = 1$.

Diszkrét valószínűségi eloszlás: Legyen ξ olyan diszkrét valószínűségi változó, melynek értékkészlete $x_1, x_2, ...$ Ekkor az $A_i, i = 1, 2, ...,$ halmazok teljes eseményrendszert alkotnak. Ebből következik, hogy a

$$p_i = P(A_i) = P\{\xi = x_i\}, i = 1, 2, ...$$

számok diszkrét eloszlást alkotnak (azaz $p_i \geq 0$, és $\sum_{i=1}^{\infty} p_i = 1$.).



Például: Két dobókockával dobunk. Az x tengelyen a dobott számok összege szerepel, az y tengelyen pedig az egyes értékek valószínűsége. Minden valószínűség nagyobb, mint nulla, és az összegük egy, mivel az események teljes eseményrendszert alkotnak. ξ csak véges, vagy megszámlálhatóan végtelen számértékeket vehet fel (2, 3, 4, ... 12).

Binomiális eloszlás: Ha egy kísérletet n-szer függetlenül megismételünk, és ξ jelenti a p valószínűségű A esemény bekövetkezéseinek a számát, akkor

$$P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, ..., n.$$

Például: visszatevéses húzás.

Poisson eloszlás: Azt mondjuk, hogy ξ λ paraméterű *Poisson-eloszlású*, ha

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, ...,$$

ahol $\lambda > 0$ konstans.

Folytonos valószínűségi eloszlás: ξ valószínűségi változó eloszlása folytonos pontosan akkor, ha az eloszlásfüggvénye folytonos függvény.

Legyen (Ω, F, P) egy valószínűségi mező. A $\xi : \Omega \to \mathbb{R}$ leképezést **folytonos valószínűségi változó**nak nevezzük, ha bármely rögzített $x \in \mathbb{R}$ esetén

$$\{\omega \in \Omega : \xi(\omega) < x\} \in F.$$

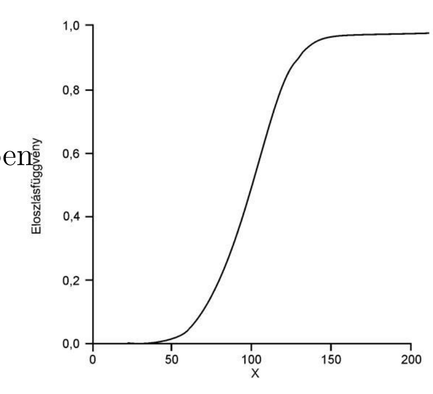
Egy $\xi : \Omega \to \mathbb{R}$ valószínűségi változó **eloszlásfüggvény**én azt az $F_\xi(x)$ függvényt értjük, melyre teljesül

$$F_\xi(x) = P(\{\omega \in \Omega : \xi(\omega) < x\})$$

Ez a függvény

- monoton nemcsökkenő
- balról folytonos
- végtelenben vett határértéke egy, mínusz végtelenben vett határértéke nulla.

Magyarán ez a függvény azt adja meg, hogy mennyi a valószínűsége annak, hogy $\xi < x$.



4.1. Az egyenletes eloszlás

4.1.1. Az egyenletes eloszlás jelentése Ha egy véges intervallumra úgy dobunk egy pontot, hogy az intervallum bármely részintervallumára annak hosszával arányos valószínűséggel essen, akkor a pont x-koordinátája egyenletes eloszlású.
A ξ valószínűségi változót az $[a, b]$ intervallumon *egyenletes eloszlásúnak* nevezzük, ha eloszlásfüggvénye

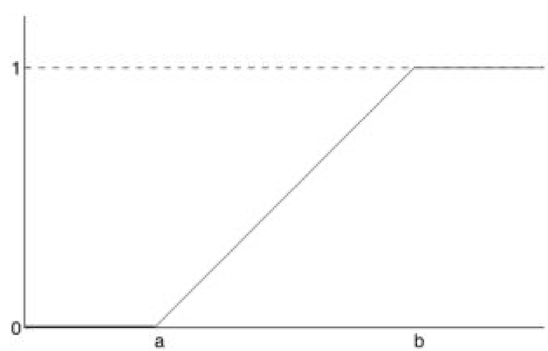
$$F(x) = \begin{cases} 0, & \text{ha} \quad x \leq a, \\ \frac{x-a}{b-a}, & \text{ha} \quad a < x \leq b, \\ 1, & \text{ha} \quad b < x. \end{cases}$$

Az egyenletes eloszlás sűrűségfüggvénye

$$f(x) = \frac{1}{b-a}, \quad \text{ha} \quad a \leq x \leq b,$$

egyébként $f(x) = 0$.

4.1.1. ábra. Az egyenletes eloszlás eloszlásfüggvénye



4.1.2. Az egyenletes eloszlás jellemző mennyiségei A várható érték:

$$\mathbb{E}\xi = \frac{a+b}{2}.$$

A szórásnégyzet:

$$\mathbb{D}^2\xi = \frac{(b-a)^2}{12}.$$

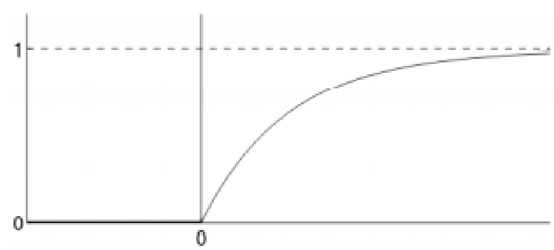
4.2. Az exponenciális eloszlás

4.2.1. Az exponenciális eloszlás definíciója A ξ valószínűségi változót λ paraméterű *exponenciális eloszlásúnak* nevezzük, ha eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Itt $\lambda > 0$ rögzített.

4.2.1. ábra. Az exponenciális eloszlásfüggvény



Az exponenciális eloszlás élettartamok és várakozási idők eloszlásaként lép fel. Az exponenciális eloszlás és a vele kapcsolatos más eloszlások a sorbanállás-elméletben és a megbízhatóság-elméletben használatosak.

Az exponenciális eloszlás sűrűségfüggvénye:

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

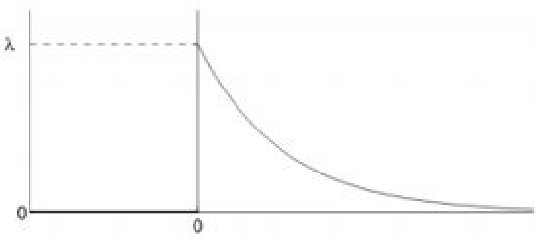
4.2.2. Az exponenciális eloszlás jellemző mennyiségei A momentumok:

$$\mathbb{E}\xi^k = \frac{k!}{\lambda^k}, \qquad k = 1, 2, \ldots.$$

Speciálisan, a várható érték és a szórásnégyzet:

$$\mathbb{E}\xi = \frac{1}{\lambda}, \quad \mathbb{D}^2\xi = \frac{1}{\lambda^2}.$$

4.2.2. ábra. Az exponenciális sűrűségfüggvény



4.2.3. Az exponenciális eloszlás tulajdonságai Az exponenciális eloszlás „örökifjú":

$$P(\xi < t + s | \xi \geq t) = P(\xi < s), \qquad t > 0,\ s > 0.$$

A fenti egyenlőség jellemzi is az exponenciális eloszlást az abszolút folytonos eloszlások között.

4.3. A normális eloszlás

4.3.1. A normális eloszlás definíciója A normális eloszláson alapul a statisztika klasszikus elméletének túlnyomó része. A ξ valószínűségi változót *normális eloszlásúnak* nevezzük, ha sűrűségfüggvénye:

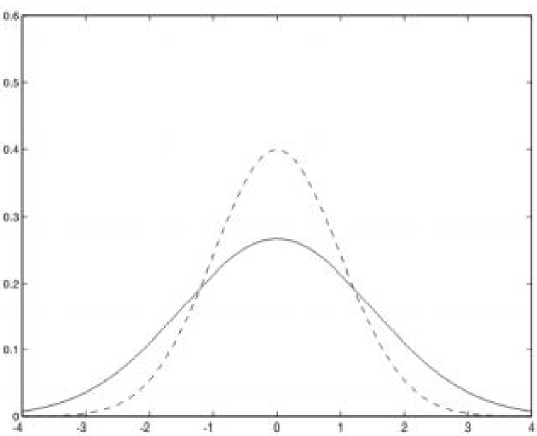
$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right)}, \tag{4.3.1}$$

ahol $m \in \mathbb{R}$, $\sigma > 0$.

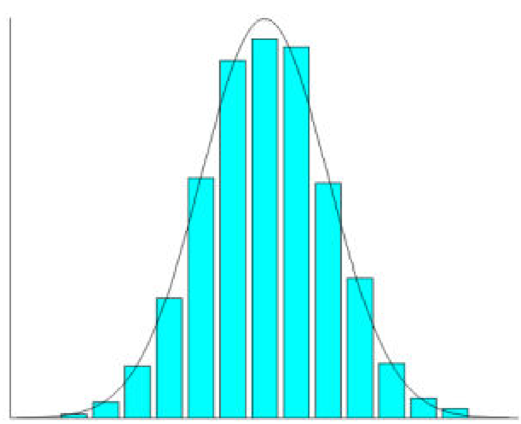
Jelölése: $\xi \sim \mathcal{N}(m, \sigma^2)$. Igazolnunk kell, hogy (4.3.1) valóban sűrűségfüggvényt határoz meg.

f grafikonja az ún. haranggörbe (Gauss-görbe). Az f függvény m-re szimmetrikus, f szigorúan monoton növekvő a $(-\infty, m]$ intervallumon. $m \pm \sigma$-ban f-nek inflexiós pontja van. m-ben f-nek maximumhelye van, a maximum értéke $\frac{1}{\sqrt{2\pi}\sigma}$. σ növelésével a harang alakú görbe „laposabbá" válik, σ csökkentésével pedig „csúcsosabbá". A 4.3.1. ábrán normális sűrűségfüggvények láthatóak $m = 0$ esetén. A csúcsosabbnál $\sigma = 0.7$, a folytonos vonallal ábrázoltnál $\sigma = 1$.

4.3.1. ábra. Normális sűrűségfüggvények különböző szórásokra

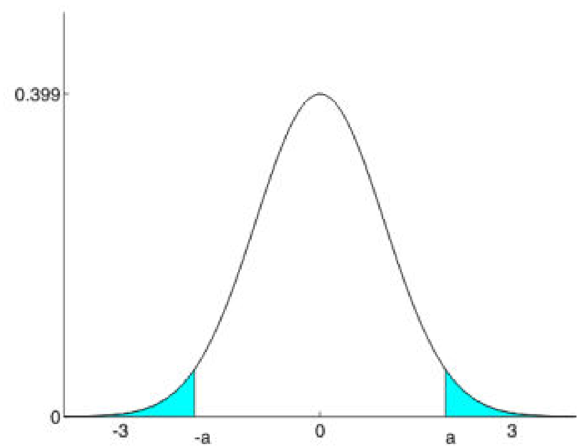


4.3.2. ábra. Hisztogram és normális sűrűségfüggvény

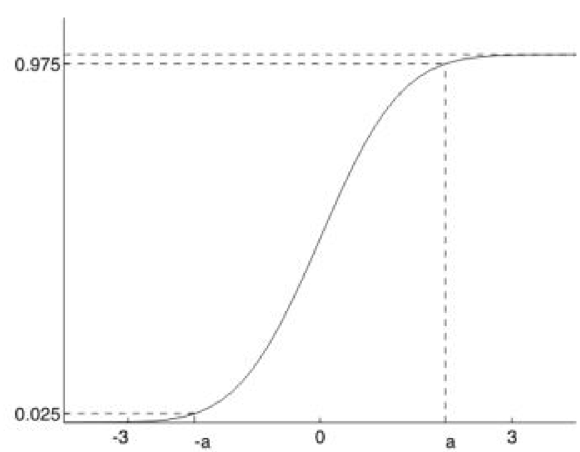


4.3.2. A standard normális eloszlás Ha $\xi_0 \sim \mathcal{N}(0,1)$, akkor ξ_0-at *standard normális eloszlásúnak* nevezzük. A 4.3.2 és a 4.3.2 ábrán a standard normális sűrűségfüggvény, ill. eloszlásfüggvény látható. Az ábrákon bejelöltük a 0.025 kvantilist: $-a = -1.96$ és a 0.975 kvantilist: $a = 1.96$. Ez azt jelenti, hogy a sűrűségfüggvény alatt besatírozott két rész mindegyike 0.025 területű. Továbbá, hogy az eloszlásfüggvény értéke a $-a = -1.96$ helyen 0.025, az $a = 1.96$ helyen pedig 0.975.

4.3.3. ábra. A standard normális sűrűségfüggvény



4.3.4. ábra. A standard normális eloszlásfüggvény



2 Második rész

Adatszerkezetekkel kapcsolatos alapfogalmak: absztrakció (logikai és fizikai szint), absztrakt adatszerkezetek (homogén-heterogén, statikus-dinamikus, struktúra, műveletek). Elemi adatszerkezetek: lista, verem, sor. Halmaz, multihalmaz, mátrix. Fák ábrázolása, keresések, bejárások, törlés, beszúrás.

Az adatelemek lehetnek egyszerűek (atomiak) vagy összetettek. Minden adatelem rendelkezik valamilyen értékkel. Az adatelemek között jól meghatározott kapcsolatrendszer van. Az adatelemek és a közöttük lévő kapcsolatok definiálják a **logikai** (absztrakt) adatszerkezetet. Független hardvertől, szoftvertől.

Fizikai adatszerkezet (társzerkezet): adatszerkezet az operatív tárban vagy periférián (háttértáron).

Absztrakt adatszerkezetek osztályozása
Lehetséges csoportosítási szempontok:

- Változhat-e az adatszerkezet elemeinek a száma?
  - statikus (nem változhat)
  - dinamikus (változhat)
- Milyen az adatszerkezet elemeinek a típusa?
  - homogén (ugyanolyan típusú minden elem)
  - heterogén
- Milyen kapcsolatban állnak egymással az adatelemek az adatszerkezetben? Egy homogén adatszerkezet lehet:
  - struktúra nélküli
  - asszociatív
  - szekvenciális
  - hierarchikus
  - hálós

  A heterogén adatszerkezeteket nem csoportosítjuk ilyen szempont alapján.

Absztrakt adatszerkezetekkel végezhető műveletek (alapvető algoritmusok):

- Létrehozás
- Módosítás
  - bővítés
  - törlés (fizikai, logikai)
  - csere
- Rendezés
- Keresés
- Elérés
- Bejárás
- Feldolgozás

Lista

– Folytonos reprezentáció: vektorral

– Szétszórt reprezentáció: láncolt listával: Egy tárhelyen egy adatelem értékét (adatrész) és legalább egy mutató értékét (mutatórész) tároljuk. A mutatók értékei memóriacímek lehetnek, amelyek megmondják az adatelem rákövetkezőinek tárbeli helyét. A tárhelyek mérete nem szükségképpen azonos, elhelyezkedésük a memóriában tetszőleges.

– Szétszórt reprezentáció alapvető fajtái:

> egyirányban láncolt lista: A tárhely (listaelem) az adatelem értékén kívül egy mutatót tartalmaz, amely a következő listaelem címét tartalmazza.
>
> cirkuláris lista: Hasonló az egyirányban láncolt listához, ám itt egyik listaelem mutatórésze sem tartalmazhatja a NIL értéket: az utolsó listaelem mutatórészébe az első listaelem címe kerül.

kétirányban láncolt lista: Hasonló az egyirányban láncolt listához, ám itt minden listaelem mutatórésze két részből áll: az egyik mutató az adott listaelemet megelőző, a másik az adott listaelemet követő listaelemre mutat.

multilista: Több változata is van. A listaelemek adatrésze összetett. Pl: Az adatrész minden komponensére fölépíthető egy egyirányban láncolt lista.

Verem: Speciális lista adatszerkezet. Csak a verem tetejére lehet betenni, illetve csak onnan lehet kivenni (LIFO). Az utoljára betett elem a verem tetejére kerül. Az elsőnek betett elem a verem aljára kerül.

Sor: Speciális lista adatszerkezet, melynek alapműveletei a speciális listaműveletek közül a következők:

– az első elemhez történő hozzáférés: ACCESS HEAD

– bővítés az utolsó elem mögé: PUT (INJECT)

– az első elem törlése: GET (POP)

FIFO, az első elemet dolgozzuk fel először, és a sor végéhez fűzzük az új elemeket.

Halmaz és multihalmaz: A halmaz és a multihalmaz struktúra nélküli, homogén és dinamikus adatszerkezetek. A halmaz minden eleme különböző. A multihalmazban előfordulhatnak azonos elemek is. Mindkét adatszerkezetre igaz, hogy az adatszerkezetben lévő elemek között nincs kapcsolat (ezért struktúra nélküli adatszerkezetek). A halmaz adatszerkezet a matematikai halmaz fogalom megjelenése az adatszerkezetek szintjén. Mindig véges – ennyiben nem felel meg teljesen a matematikai halmaz fogalmának. Alapműveletek: eleme, unió, metszet, különbség.

Klasszikus reprezentációja folytonosan, karakterisztikus függvény segítségével történik: a lehetséges elemeket sorba rendezzük, és mindegyikhez egy bit méretű tárterületet rendelünk: 1 ha benne van az adott elem a halmazban, 0 ha nem (multihalmaznál több területet, az ott lévő érték azt mondja meg, hányszor szerepel az elem a multihalmazban).

Mátrix: Kettő (vagy több) dimenziós tömb. Statikus, homogén és asszociatív adatszerkezet. A felépítése definiálja: benne az adatelemek egymáshoz viszonyított helyzete a lényeges. A tömb bármelyik eleme egész számok sorozatán keresztül érhető el. Minden adatelemhez különböző egészszámsorozat tartozik, így az asszociativitást biztosító részhalmazok egyeleműek és diszjunktak. A számsorozat számait indexeknek nevezzük, segítségükkel tudjuk az adatelemet kiválasztani. Az indexek darabszámát a tömb dimenziójának hívjuk. Folytonos reprezentációnál a leképezés lehet sorfolytonos vagy oszlopfolytonos.

Felsőháromszög-mátrix: olyan kvadratikus (négyzetes) mátrix, melynek főátlója alatt csupa 0 elem található.

Alsóháromszög-mátrix: olyan kvadratikus (négyzetes) mátrix, melynek főátlója felett csupa 0 elem található.

Szimmetrikus mátrix: olyan kvadratikus (négyzetes) mátrix, melynél $a_{i,j} = a_{j,i}$

Ritka mátrix: A ritka mátrixok olyan (általában nagyméretű) mátrixok, amelyekben a legtöbb elem értéke ugyanaz (általában 0). Az ettől eltérő értékkel rendelkező elemeket ritka elemeknek nevezzük. 3 soros reprezentáció: sor, oszlop, érték.

Fa: Homogén, dinamikus, hierarchikus adatszerkezet. Fa adatszerkezetekkel kapcsolatos fogalmak:

- csúcs, csomópont
- gyökérelem
- levélelem
- közbenső elem
- él
- út
- részfa
- szint
- magasság

Rendezetlen fáknál nem lényeges az ugyanazon csúcsból kiinduló élek sorrendje, rendezett fáknál viszont igen.

Bináris fa: Olyan fa, melyben minden adatelemnek legfeljebb két rákövetkezője van. Bináris fa bejárásai:

- *Preorder:* Gyökér után bal oldali részfa preorder módon, majd a jobb oldali preorder módon
- *Inorder:* Bal, gyökér, jobb
- *Postorder:* Bal, jobb, gyökér

Kupac: A kupac olyan fa, amely rendelkezik a kupac tulajdonsággal: a gyökérelemet kivéve bármely adateleménének a kulcsa kisebb vagy egyenlő az adatelem szülőjének a kulcsánál.

Minimális magasságú bináris fa: Azt mondjuk, hogy egy bináris fa minimális magasságú, ha adott számú elemet nem lehetne kisebb magasságú bináris fában elhelyezni.

Tökéletesen kiegyensúlyozott bináris fa: Azt mondjuk, hogy egy bináris fa tökéletesen kiegyensúlyozott, ha bármely elemének bal és jobb oldali részfájában az elemek darabszáma legfeljebb 1-gyel tér el.

Bináris keresőfa: A bináris keresőfa olyan rendezett bináris fa, melyben az adatelemek mindegyike rendelkezik egy kulccsal, és minden adatelemre igaz az, hogy az adatelem bal oldali részfájában lévő elemek kulcsai kisebbek, a jobb oldali részfájában lévő elemek kulcsai pedig nagyobbak az elem kulcsánál. Bővítése rekurzívan: ha üres a fa, a bővítendő elem lesz az egyetlen levélelem, az algoritmus véget ér. Ha a beszúrandó elem kisebb a gyökérelemnél, akkor a gyökérelem bal oldali részfáját bővítjük a beszúrandó elemmel. Egyébként a gyökérelem jobb oldali részfáját bővítjük a beszúrandó elemmel.

Törlés bináris keresőfából rekurzívan

1. Ha üres a fa, akkor nem tudunk törölni, és ezzel az algoritmus sikertelenül véget ér.
2. Összehasonlítjuk a gyökérelem értékét a törlendő elemmel.
   1. Ha a törlendő elem kisebb a gyökérelemnél, akkor a gyökérelem bal oldali részfájából töröljük a törlendő elemet.
   2. Ha a törlendő elem nagyobb a gyökérelemnél, akkor a gyökérelem jobb oldali részfájából töröljük a törlendő elemet.
   3. Ha a két elem egyenlő, akkor megnézzük, hogy a gyökérelemnek hány rákövetkezője van.
      i. Ha a gyökérelemnek egy rákövetkezője sincs (azaz levélelem), akkor egyszerűen törölhető.
      ii. Ha a gyökérelemnek egy rákövetkezője van, akkor felülírjuk a gyökérelemet azzal a rákövetkező elemmel (azaz egy szinttel feljebb csúsztatjuk a gyökérelem nem üres részfáját).
      iii. Ha a gyökérelemnek két rákövetkezője van, akkor a gyökérelem értékét felülírjuk a gyökérelem **jobb oldali részfája legbaloldalibb elemének** az értékével, majd a gyökérelem jobb oldali részfájából töröljük ezt a legbaloldalibb elemet.

   Ezzel az algoritmus sikeresen véget ér.

Kiegyensúlyozott bináris fa

Azt mondjuk, hogy egy bináris fa kiegyensúlyozott, ha bármely elemére igaz, hogy az elem bal oldali és jobb oldali részfájának magasságkülönbsége legfeljebb 1.

Megjegyzés

Minden tökéletesen kiegyensúlyozott fa egyben kiegyensúlyozott is.

Kiegyensúlyozott keresőfa (AVL-fa)

Akkor nevezünk egy bináris fát kiegyensúlyozott keresőfának vagy AVL-fának, ha kiegyensúlyozott is és keresőfa is egyben.

Piros-fekete fa

A piros-fekete fa olyan bináris keresőfa, amely a következő tulajdonságokkal rendelkezik:

1. Minden csomópontja piros vagy fekete.
2. A gyökere fekete.
3. Minden (NIL értékű) levele fekete.
4. Ha egy csomópont piros, akkor mindkét rákövetkezője fekete. (Más szavakkal kifejezve: nincs benne két egymást követő piros csomópont.)
5. Minden csomópont esetén az összes olyan úton, amely az adott csomópontból indul ki és levélig vezet, ugyanannyi a fekete csomópontok száma. (Beszélhetünk fekete magasságról.)