Valószínűségszámítás és statisztika
4. gyakorlat
2014. 10. 02.

Elmélet

[Minta] $X_1, ..., X_n$ valószínűségi változó sorozat. (Jel. $\mathbf{X} = X_1, ..., X_n$)
A továbbiakban feltesszük, hogy függetlenek és azonos eloszlásúak. Magyarosan rövidítve FAE minta, de gyakrabban használják az angol *i.i.d. minta* rövidítést (independent, identically distributed).
Az elméleti értékeket nagy, a konkrét, realizált mintából számolt értékeket mindig kis betű fogja jelölni, azaz minta esetén $x_1, ..., x_n$.

[Statisztika] A minta valamely függvénye: $T : \mathbf{X} \to ...$

[Becslés] A minta eloszlásának ismeretlen paraméterét közelíti a minta segítségével. Megj.: Minden becslés statisztika.

Néhány lényeges statisztika:

- *Rendezett minta*: $X_1^* \leq ... \leq X_n^*$ nem csökkenő sorrendbe tesszük a mintaelemeket.
- *Mintaátlag*: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- *Tapasztalati szórás*: $S_n = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}$
  Értelmezése: az átlagtól való átlagos eltérés abszolút mértékegységben.
- *Korrigált tapasztalati szórás*: $S_n^* = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$
- *Szórási együttható*: $V = \frac{S_n}{\overline{X}}$
  Értelmezése: az átlagtól való átlagos eltérés százalékban. Megj.: relatív szórásnak is hívják.
- *Tapasztalati eloszlásfüggvény*: $F_n(x) = \frac{\sum_{i=1}^{n} I(X_i < x)}{n}$
  ahol $I(X_i < x) = \begin{cases} 1 & \text{ha } X_i < x \\ 0 & \text{ha } X_i \geq x \end{cases} \rightsquigarrow$ indikátor függvény
- *z-kvantilis*: $q_z = \inf\{x : F(x) > z\}$, és amennyiben F invertálható, akkor $q_z = F^{-1}(z)$-re egyszerűsödik.
  Értelmezése: a mintaelemek z-ed része q_z-nél kisebb, $(1 - z)$-ed része q_z-nél nagyobb
  Realizált mintából sokféleképpen számolható, interpolációs módszer:
  1.) Sorszám megállapítása: $(n + 1)z = e + t$ (e:egészrész, t:törtrész)
  2.) $q_z = X_e^* + t(X_{e+1}^* - X_e^*)$
- *kvartilisek*: speciális kvantilisek
  - $Q_1 := q_{\frac{1}{4}} \rightsquigarrow$ alsó kvartilis
  - $Q_2 = Me := q_{\frac{1}{2}} \rightsquigarrow$ medián (középső mintaelem)
  - $Q_3 := q_{\frac{3}{4}} \rightsquigarrow$ felső kvartilis

Személyes adatok:
Név: Bondici László
E-mail: bondici@cs.elte.hu
Honlap: www.cs.elte.hu/∼bondici

Feladatok

1. (3-as feladatsor 9-es feladat) A sarki boltban vásárlásunk után újságot kapunk ajándékba. 10 fajta ilyen újság van, mindegyik egyformán gyakori. Mennyi a valószínűsége, hogy a 20. vásárlásnál lesz meg mind a 10 újságunk?

2. Tegyük fel, hogy két cégnél az alkalmazottak havi jövedelme (eHUF-ban) az alábbiak szerint alakult:
   Süti Sütőde (10 dolgozó): 60, 60, 60, 90, 100, 100, 120, 140, 170, 500.
   SST szoftvercég (50 dolgozó): 6 fő: 60, 10 fő: 100, 10 fő: 150, 10 fő: 200, 10 fő: 500, 3 fő: 600, 1 fő:1000.
   Ábrázoljuk az adatokat hisztogrammal, számoljuk ki a különböző középértékeket (átlag, medián). Arra az esetre is számoljunk, ha a kiugróan magas 500 (1000) eHUF-os értékekről rájövünk, hogy csak elírás, és valójában 50 eHUF, illetve 100 eHUF a helyes.

3. Tegyük fel, hogy egy egyetemen az alábbi felvételi arányok adódtak:

Kar	Összes jelentkező (ebből fiú)	Felvettek (ebből fiú)
Gépészmérnöki	202 (180)	170 (150)
Jogi	598 (320)	130 (50)

   Ábrázoljuk a karonkénti adatokat kontingenciatáblázat formájában. Mit mondhatunk a fiúk, illetve a lányok felvételi arányáról az egyes karokon és összesítve?

4. Tekintsük a KSH által közzétett éves fogyasztói árindex-adatokat:

Év	Árindex	Év	Árindex	Év	Árindex	Év	Árindex
1998	114,3	2001	109,2	2004	106,8	2007	108,0
1999	110,0	2002	105,3	2005	103,6	2008	106,1
2000	109,8	2003	104,7	2006	103,9	2009	104,2

   Ábrázoljuk az adatokat hisztogram és boxplot segítségével. Számoljuk ki a mintaátlagot és a mediánt.

5. Tegyük fel, hogy egy cégnél két egymás utáni évben az alábbi fizetés-eloszlások adódtak:

Havi fizetés (eHUF, 2004)	Gyakoriság	Havi fizetés (eHUF, 2005)	Gyakoriság
50-79	13	50-79	13
80-119	10	80-119	17
120-159	7	120-159	16
160-240	2	160-240	4

   Ábrázoljuk az adatokat kördiagrammal.

6. Dobjunk fel 6-szor egy kockát.
   a) Határozzuk meg a középértékeket: átlag, medián.
   b) Adjuk meg a rendezett mintát.
   c) Határozzuk meg minden valós z-re annak a relatív gyakoriságát, hogy a kísérletünk során z-nél kisebb értéket kaptunk. Ábrázoljuk is a kapott függvényt.

7. Iszákos Iván a nap 2/3 részét kocsmában tölti. Mivel a faluban 5 kocsma van, és nem válogatós, azonos eséllyel tartózkodik bármelyikben. Egyszer elindulunk, hogy megkeressük. Négy kocsmát már végigjártunk, de nem találtuk. Mi a valószínűsége annak, hogy az ötödikben ott lesz?

8. Egy szabálytalan érmével dobunk (p a fej valószínűsége). Jelölje X az első, azonosakból álló sorozat hosszát (Például, ha a sorozat FFIII…, akkor X=2.) Adjuk meg X eloszlását!

Házi feladatok

1. Szimuláljunk adatokat visszatevéses és visszatevés nélküli mintavétellel, ugyanabból a sokaságból, különböző esetekre. Ábrázoljuk és értékeljük az eredményeket.