

### Kiindulás:

Egy Gábor által megadott konkrét diszkrét valószínűségi mérték mely megfelel egy 7-dimenziós valószínűségi vektorváltozó eloszlásának, mely minden koordinátájában az 1; 2; 3; 4; 5 értékeket veheti fel.

Az eloszlás a relo.dat fileban van megadva. Jelöljük  $PR$ -el. Gábor tájékoztatása szerint a koordináták függetlenségére vett  $\pi^*$ -ja  $\frac{1}{4}$ .

### Vágás iterációs művelet:

Bemenet: az előző 7-dimenziós alaphalmazon vett  $Q$  mérték és egy  $v$  pozitív valós szám.

0. lépés: Véletlen permeloszlások konstruálása. Például 5 db.  $U(0,1)$ -et generálunk, utána visszanzormáljuk őket, hogy összegük 1 legyen, ezt 7-szer megismételjük. Ebből meghatározzuk a szorzatmértéket. Tapasztalat szerint a végső eredmény nem függ a kezdeti ponttól.
1. iterációs lépés: Vesszük a  $P$  szorzatmértéket,  $v$ -t és  $Q$ -t.  $P_1 := \min(vP, Q)$
2. iterációs lépés:  $t := \text{sum}(P_1)$
3. iterációs lépés:  $P_2 := P_1/t$
4. iterációs lépés:  $P_3 := P_2$  marginálisainak szorzatmértéke
5. lépés: visszatérünk az 1. iterációs lépéshez  $P_3$ -al, mint  $P$ -vel.

Leállás: amikor a két egymás utáni szorzatmérték pontonkénti abszolút eltéréseinek összege nem haladja meg  $10^{-8}$ -t.

Kimenet:

az utolsó  $t$ :  $t(v)$

az utolsó  $P_2$  és  $P_3$  Kullback–Leibler divergenciája:  $D(v) := \sum_i P_2(i) \log \left( \frac{P_2(i)}{P_3(i)} \right)$

az utolsó  $P_3$  és az eredeti  $PR$  négyzetes eltérése (négyzetes eltérések összege)

minden cellában jelzés arra, hogy az utolsó  $P_2$  vagy  $Q$  volt a kisebb:  
 $\begin{cases} 1, & \text{ha } Q(i) \leq P_2(i) \\ 0, & \text{különben} \end{cases}$

### PR vizsgálata:

Először 0,51-től 1,5-ig végeztük el a vágásokat. Az eredmények ( $v$ ,  $t(v)$  és  $D(v)$ ) a vagas.csv fileban vannak. Gábor várakozásának megfelelően  $v=0,75$ -re kaptuk  $t(0,75)=0,75$ -öt.

A 0,75-ös vágásra a részletes eredmények az eredeti.csv fileban vannak. Itt látható, hogy a kapott szorzatmérték és az eredeti mérték négyzetes eltérése 1,4%. A vágás utolsó lépésében az eredeti *PR* mérték 27652-szer volt a kisebb és 50473-szor a nagyobb.

### **Minták vizsgálata:**

Megnéztük, hogy különböző mintaelemszámnál mit adnak a tapasztalati eloszlások. A mintaelemszámok 100 E, 1 M, 2 M, 5 M és 10 M voltak. Mindegyik esetben 100-szor ismételtük meg a kísérletet. Meghatároztuk a tapasztalati eloszlásokat és mindig azt a vágást kerestük meg, amire  $t(v)$  0,75-ös értéket adott.

Az eredmények az ered"mintaelemszám".RData file-okban vannak (ezeket egy kicsit később töltöm fel, egy másik gépen vannak):

1. oszlop: mintaelemszám
2. oszlop: kísérlet sorszáma
3. oszlop: vágás ( $v$ )
4. oszlop:  $t(v)$
5. oszlop: mintaelemszám\*divergencia
6. oszlop: a független marginálisokból képzett mérték négyzetes eltérése a *PR* mértéktől
7. – 78131. oszlop: a tapasztalati eloszlás adja-e minimumot a vágásnál

Az ered"mintaelemszám".csv file-ok tartalmazzák az első 6 oszlopot.

A Qgyak"mintaelemszám".csv fileok tartalmazzák azokat a számokat, hogy hány cella esetében volt kisebb a vágásnál a tapasztalati eloszlás.

Ezeknek az értékeknek a hisztogramja látható a Qgyak"mintaelemszám".jpeg képeken.

### **R program:**

picsillag20160902.R