Véletlen partíciók generálása

Vincze Erika

2011. április 1.

Tartalomjegyzék

1. Bevezetés

A dolgozatban a véletlen partíciók generálásával foglalkozom. Ehhez $[n] = \{1, 2, \ldots\}$ egyenletes véletlen permutációját használom fel. Ez a kétparaméteres felcserélhető véletlen partíciók családjának vizsgálatához vezet, melyek különböző módokon jellemezhetőek, és szoros kapcsolatban állnak a gamma és stable subordintarokhoz(?).

A kínai étterem reprezentáció

Ez a folyamat egy σ_n véletlen permutációt definiál az $[n] := \{1, ..., n\}$ halmazon úgy, hogy a Π_n véletlen partíciókat σ_n ciklusai generálják. Így a pozitív egész számok legáltalánosabb véletlen partícióihoz jutunk.

A kétparaméteres modell

Ez a bekezdés legfőbbképp a természetes számok véletlen partícióval foglalkozik, melyet valós számok egy (α, θ) párjával paraméterezünk. Ezzel a $GEM(\alpha, \theta)$(Griffith-Engen-McCloskey), illetve sorba rendezett esetben a $PD(\alpha, \theta)$ (Poisson-Dirichlet) eloszlásokhoz jutunk.

2. A kínai étterem reprezentáció

Konzisztens véletlen permutációk
Tekintsük a $(\sigma_n, n = 1, 2, ...)$ véletlen permutációk egy sorozatát a következőképpen:

1. σ_n egyenletes véletlen permutációja [n]-nek, minden n-re.

2. Ha σ_n adott a ciklusokból minden n-re, akkor σ_{n-1}-et kapjuk, ha σ_n ciklusából töröljük az n. elemet.

Például, a sztenderd jelöléseket használva:
ha $\sigma_5 = (134)(25)$ akkor $\sigma_4 = (134)(2)$;
ha $\sigma_5 = (134)(2)(5)$ akkor $\sigma_4 = (134)(2)$.
Könnyen belátható, hogy a fenti eljárással a következő eljárásnak megfelelő eloszlású σ_n sorozathoz jutunk:
Egy kezdetben üres étteremben korlátlan számú kerek asztal áll rendelkezésre, 1; 2;... számozással, és minden egyes asztalhoz korlátlan számú ember ülhet. A vendégek is 1; 2;.... számozásúak, és egyenként foglalnak helyet a következők szerint:

Egyszerű véletlen ülésrend

Az első ember az első asztalhoz ül. Az n+1-edikhez tegyük fel, hogy már n ember helyet foglalt valamelyik asztalnál úgy, hogy minden j-re $(1 \leq j \leq k)$, ahol k az olyna asztalok száma, amelynél az első n ember ül. Az n+1-edik vendég a következő n+1 hely közül egyenlő valószínűséggel választ: a j. vendég bal oldalára $(1 \leq j \leq n)$, vagy egyedül ül a k+1-dik asztalhoz.
Legyen $\sigma_n : [n] \rightarrow [n]$ a következő: ha az n. vendég érkezése után az i. és a j. vendég egy asztalnál ül, i. a j. bal oldalán, akkor $\sigma_n(i) = j$, és ha az i. vendég egyedül ül egy asztalnál, akkor $\sigma_n(i) = i$.
σ_n ekkor kielégíti a (1) ill. (2) feltételeket.
Ebből a konstrukcióból az egyenletes véletlen permutációk sok tulajdonsága kiolvasható.
Az n. vendég érkezése után az asztaloknál ülő emberek száma a következőképpen alakul:

$$K_n = \{\sigma_n ciklusai\} = Z_1 + ... + Z_n$$

ahol Z_j az indikátora annak az eseménynek, hogy a j. vendég új asztalhoz ül. A konstrukció miatt a Z_j-k független Bernuolli(1/j) valószínűségi változók, ezért:
$\frac{K_n}{\log n} \rightarrow 1$ majdnem biztosan, és $\frac{K_n - \log n}{(\log n)^{1/2}} \xrightarrow{d} B_1$,
ahol B_1 a standard Gauss eloszlás.
Legyen Π_n a σ_n ciklusai által generált véletlen partíciója [n]-nek. Ekkor Π_n az [n] felcserélhető véletlen partíciója, és Π_n konzisztens, ahogy n változik. Így a $\Pi_\infty = (\Pi_n)$ sorozat az N egy felcserélhető véletlen partíciója. Legyen X_n annak az eseménynek az indikátora, hogy az n+1. vendég az első asztalhoz ül. Ekkor az $(X_n)_{n \geq 1}$ egy felcserélhető sorozat, ami a Pólya-féle urnamodellnek az eredménye, a = b = 1-re. Ezért $S_n = X_1 + ... + X_n$ egyenletes eloszlású

a {0; 1;...;n}-on. Ugyanígy az $S_n + 1$ méretű ciklusa az 1-et tartalmazó σ_{n+1}-nek egyenletes eloszlású a {1;...;n+1}-en. Az aszimptotikus gyakorisága a σ_n 1-et tartalmazó osztályainak, majdnem biztosan korlátos, $S_n = n$-nel, ami természetesen egyenletes eloszlású a [0; 1]-en.

A határeloszlások

Legyen $(N_{n,1}; ...; N_{n,K_n})$ a Π_n blokkjainak mérete, az elemek megjelenése szerinti sorrendben. Az éttermes modell szerint $N_{i,n}$ az i. asztalnál ülök száma, amikor n ember van az étteremben. Az előzőek szerint $N_{n,1}$ egyenletes eloszlású $[n]$-en. Hasonlóan, adott $N_{n,1} = n_1 < n$, esetén $N_{n,2}$ is egyenletes eloszlású $[n - n_1]$-en, és így tovább. Ez a diszkrét egyenletes "stick-breaking" séma, melynek eloszlása, ha $n \to \infty$, a relatív gyakorisága σ_n ciklusainak a méretének $(N_{n,i}/n; i \geq 1)$ tart a folyamatos egyenletes "stick-breaking" folyamathoz eloszlásban.

$$(P_1, P_2, \ldots) = (U_1, U_1^- U_2, \ldots),$$

ahol U_i-k független $U[0, 1]$ változók, és $U^- = 1 - U$. Ekkor a hozzá tartozó EPPF:

$$p_{0,1}(n_1, \ldots, n_k) := \frac{1}{n!} \prod_{i=1}^{k} (n_i - 1)!.$$

Általánosítás:

A kínai étterem folyamat könnyen általánostíható, ha a σ_n permutációk sorozatához tartozó $\pi_\infty := (\pi_n)$ a legáltalánosabb felcserélhető partíciója az egész számoknak. A hozzá tartozó EPPF: $p(n_1, ..., n_k)$ megadja minden $(n_1, ..., n_k)$-ra annak a valószínűségét, hogy π_n az $[n]$-et éppen $(n_1, ..., n_k)$ méretű halmazokra osztja. A kínai éttermes meggondolásból most csak a σ_n ciklusok által generált partíciókkal foglalkozunk. Így $1 \leq i \leq K_n$-re az állítás, miszerint az n+1-edik látogató az i. asztalhoz ül, azt jelenti, hogy π_{n+1} az $[n + 1]$ olyan partíciója, melynek $[n]$-re való megszorítása π_n, és n+1 a π_n i. osztályához tartozik. Hasonlóan, ha az n+1. vendég új asztalhoz ül, akkor nk. Így $1 \leq i \leq K_n$-re az állítás, miszerint az n+1-edik látogató az i. asztalhoz ül, azt jelenti, hogy π_{n+1} az $[n+1]$ olyan partíciója, melynek $[n]$-re való megszorítása π_n egy $\{n+1\}$ külön blokkal. Adott végtelen EPPF-hez tartozó felcserélhető random partíciója a természetes számoknak konstruálható a következőképpen:

Véletlen ülésrend egy felcserélhető partícióhoz:

Az első vendég az első asztalhoz ül. Ha n-ig adott π_n partíció, az első n ember helyével, k asztallal, akkor az n+1. vendég:

- a j. asztalhoz ül $p(\ldots, n_j + 1, \ldots)/p(n_1, \ldots, n_k)$ valószínűséggel
- új asztalhoz ül $p(n_1, \ldots, n_k, 1)/p(n_1, \ldots, n_k)$ valószínűséggel

. Ez megfelel a feltételeknek. Az ültetési rendek legtöbbje által generált π nem felcserélhető. Kingman elmélete szerint annak, hogy felcserélhető partíciót kapjunk, szükséges feltétele, hogy minden i-re létezik egy majdnem biztos határeloszlás az i. asztalnál ülők számára. Ennek elérésének a legegyszerűbb módja a következő:

Véletlen ülésrend egy parciálisan felcserélhető partícióhoz
Legyen $P_i, i = 1, 2, \ldots$ nem negatív véletlen változók tetszőleges sorozata úgy, hogy $\sum_{i=1}^{k} P_i \leq 1$. Adott $P_i, i = 1, 2, \ldots$ sorozathoz az első ember üljön az első asztalhoz, és ha n-ig adott az ülésrend k asztallal, akkor az n+1. ember:

- a j-dik helyre ül P_j valószínűséggel
- új asztalhoz ül $1 - \sum_{i=1}^{k} P_i$ valószínűséggel.

A konstrukció és a nagy számok törvénye miatt így léteznek a határeloszlások, melyek éppen a P_i-kel egyeznek meg. Az adott P_i sorozatra meghatározható annak a valószínűsége, hogy $[n]$ éppen $(n_1, \ldots, n_k)$ méretű blokkokra bomlik. Ez a következő formulával fejezhető ki:

$$p(n_1, \ldots, n_k) = E[\prod_{i=1}^{k} P_i^{n_i - 1} \prod_{i=1}^{k-1} (1 - \sum_{j=1}^{i} P_j)] \; (1.)$$

1. Tétel:. *Legyen (P_i) nem negatív véletlen változók sorozata úgy, hogy $\sum_{i=1}^{k} P_i \leq 1$. $p(n_1, \ldots, n_k)$-t pedig az előző formulával határozzuk meg.*

1. *$\exists\ \pi_\infty$ felcserélhető véletlen partíciója a természetes számoknak, melynek blokkjainak sűrűsége megjelenési sorrendben (P_i) eloszlása (P_i) akkor és csak akkor, ha $p(n_1, \ldots, n_k)$ függvény szimmetrikus függvénye $(n_1, \ldots, n_k)$-nek minden k-ra.*
2. *Ha π ilyan felcserélhető véletlen partíciója a természetes számoknak , akkor a π-hez tartozó EPPF: $p(n_1, \ldots, n_k)$ a (1.) alapján adódik $P_i = \mathrm{P}_i$-vel...*

3. A kétparaméteres modell

$(\alpha, \theta))$ **ültetési rend** Az n. lépésben k elfoglalt asztallal az i. asztalnál ülők száma n_i. A következő vendég

- az i. asztalhoz ül $(n_i - \alpha)/(n + \theta)$ valószínűséggel
- új asztalhoz ül $(\theta + k\alpha - \alpha)/(n + \theta)$.

A valószínűségek tulajdonságainak megfeleltetve α és θ a következők szerint választhatók.

- $\alpha = -\kappa < 0$ és $\theta = m\kappa$ valamely $m = 1, 2, \ldots$ esetén.
- vagy $0 \leq \alpha \leq 1$ és $\theta > -\alpha$.

Ekkor a következő eseteket vizsgáljuk:

($(\alpha = -\kappa < 0)$ és $\theta = m\kappa$ valamely $m = 1, 2, \ldots$ esetén) Ekkor π eloszlása m κ paraméterű szimmetrikus Dirichlet.

($\alpha = -\kappa$ és $\theta > 0$ esetén) Ez az előző eset gyenge limesze, ha $\kappa \to 0$ és $m\kappa \to \theta$. A Blackwell-MacQueen urnamodellhez hasonlóan π eloszlása θ paraméterű Dirichlet folyamatból vett mintáknak felel meg.

($\alpha = 0$ és $\theta = 1$ esetén) Az előző modellt adja.

($0 < \alpha < 1$ és $\theta > -\alpha$ esetén)