SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Vanja Jović

POISSON - DIRICHLETOVA RAZDIOBA

Diplomski rad

Voditelj rada:
Prof. dr. sc. Bojan Basrak

Zagreb, 02, 2015.

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Ovaj diplomski rad posvećen je mojoj obitelji koja me je svesrdno podržavala tijekom mojih studentskih dana te, posebno, mojoj majci koja je rezultate kolokvija iščekivala s jednakim nestrpljenjem kao i ja. Također, zahvaljujem se i svojem mentoru, prof. dr. sc. Bojanu Basraku, na svesrdnoj podršci bez koje ovaj rad ne bi ugledao svjetlo dana.

Sadržaj

Uvod

Poisson - Dirichletova razdioba, koja predstavlja vjerojatno najvažniju razdiobu na beskonačnodimenzionalnom simpleksu, prvi put je spomenuta u radu J.F.C. Kingmana 1975.g. Od tada su pronađene mnoge primjene u statistici, kombinatorici, teoriji brojeva, financijama, makroekonomiji, fizici i, posebice, populacijskoj genetici (v. [1] i [2] za listu refernci). Namjera ovog diplomskog rada jest da se čitatelj upozna sa razdiobom, sa njenim izvodima i osnovnim, odnosno graničnim rezultatima.

U prvom poglavlju želimo čitatelja upoznati sa osnovnim pojmovima koji su potrebni za razumijevanje središnjeg djela ovog rada. Započinjemo s definicijom Poissonovog procesa te iskazima i dokazima važnijih rezultata koji su vezani uz Poissonov proces. Završavamo s definicijom subordinatora i Lévyjeve mjere koji su potrebni za razmatranje rezultata vezanih uz Poisson - Dirichletov proces.

U drugom poglavlju posvećujemo se središnjoj temi rada: opisujemo kako se pronalazi razdioba te nastavljamo sa rezultatima o marginalnoj razdiobi. Rad zaključujemo sa Ewensovom formulom uzorkovanja, važnim rezultatom koji ima mnoge primjene u populacijskoj genetici.

Poglavlje 1

Osnovni pojmovi

1.1 Poissonov proces

Položaj zvijezda na nebu ne pokazuje jasan obrazac. Čini se kao da su nasumično razbacane bez očitog rasporeda ili gustoće u rasporedu. Slične situacije se mogu pojaviti u jednoj ili više dimenzija . Poissonov proces opisuje upravo takav fenomen, koristeći vjerojatnosni model koji se opisuje takvo nasumično ponašanje. Općenito, Poissonov model je obično najjednostavniji, i na neki način, *najslučajniji* način opisivanja određenog fenomena. Ime dolazi od istoimene vjerojatnosne distribucije, koja igra središnju ulogu u ovoj teoriji. Navedena distribucija je pak granični slučaj binomne distribucije i kao takva se povezuje sa radom francuskog matematičara Poissona iz 1837. godine.

Definicija 1.1.1. *Kažemo da slučajna varijabla X ima Poissonovu razdiobu ako postiže nenegativne vrijednosti i ako vrijedi*

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, k \geq 0,$$

gdje $\lambda > 0$ *zovemo parametar Poissonove distribucije.*

Poissonovu distribuciju sa parametrom λ ćemo u nastavku označavati sa $\mathscr{P}(\lambda)$. Ponekad je uputno proširiti početnu definiciju tako da se uključe i granični slučajevi 0 i ∞ definirajući u ovim slučajevima

$$\mathbb{P}(X = 0) = 1, \tag{1.1}$$

odnosno

$$\mathbb{P}(X = \infty) = 1. \tag{1.2}$$

Sada, pomoću formule za matematičko očekivanje, slijedi

$$\mathbb{E}(\mathscr{P}(\lambda)) = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \lambda,$$

pa često govorimo da je riječ o *Poissonovoj razdiobi sa očekivanjem* λ. Ako je z proizvoljan realan ili kompleksan broj sa svojstvom $|z| \leq 1$, tada je slučajna varijabla z^X ograničena te vrijedi

$$\mathbb{E}(z^X) = \sum_{k=0}^{\infty} z^k \cdot \mathbb{P}(X = k) = e^{\lambda} \sum_{k=0}^{\infty} \frac{(\lambda \cdot z)^k}{k!} = e^{-\lambda(1-z)}.$$

Jedno od najvažnijih svojstava Poissonove distribucije jest *aditivnost*: Ako su $X \sim \mathscr{P}(\lambda)$ i $Y \sim \mathscr{P}(\mu)$ dvije *nezavisne* slučajne varijable sa Poissonovm razdiobom, tada za $r, s \geq 0$ vrijedi

$$\mathbb{P}(X = r, Y = s) = \mathbb{P}(X = r)\mathbb{P}(Y = s) = \frac{\lambda^r e^{-\lambda}}{r!} \frac{\mu^s e^{-\mu}}{s!}.$$

Možemo pronaći distribuciju slučajne varijable $X + Y$ sumirajući prethodnu relaciju po r, s gdje $r + s$ držimo fiksnim:

$$\begin{aligned}
\mathbb{P}(X + Y = n) &= \sum_{r=0}^{n} \mathbb{P}(X = r, Y = n - r) \\
&= \sum_{r=0}^{n} \frac{\lambda^r e^{-\lambda}}{r!} \frac{\mu^{n-r} e^{-\mu}}{(n-r)!} \\
&= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{r=0}^{n} \binom{n}{r} \lambda^r \mu^{n-r} \\
&= \frac{(\lambda + \mu)^n e^{-(\lambda+\mu)}}{n!}
\end{aligned}$$

Znamo da funkcija izvodnica jednoznačno određuje razdiobu pojedine slučajne varijable. Prema tome, $X + Y$ ima $\mathscr{P}(\lambda + \mu)$ distribuciju. Pokazano svojstvo se može proširiti indukcijom i na zbroj proizvoljno, ali konačno mnogo, nezavisnih slučajnih varijabli, od kojih svaka ima Poissonovu razdiobu. Ipak, za naše potrebe će biti potreban i snažniji rezultat od prethodnog, koji pokriva beskonačne sume i daje uvjete za konvergenciju.

Teorem 1.1.2. *(Teorem o prebrojivoj aditivnosti)*
Neka su X_j, $j = 1, 2, ...$ nezavisne, slučajne varijable sa svojstvom $X_j \sim \mathscr{P}(\mu_j)$, za svaki $j = 1, 2, \ldots$. Ako red

$$\sigma = \sum_{j=1}^{\infty} \mu_j \tag{1.3}$$

konvergira, tada i slučajni red

$$S = \sum_{j=1}^{\infty} X_j$$

konvergira s vjerojatnošću 1 i S ima $\mathscr{P}(\sigma)$ distribuciju. Također, ako (1.3) *divergira, tada S divergira sa vjerojatnošću 1.*

Dokaz. Indukcijom po n slijedi da

$$S_n = \sum_{j=1}^{n} X_j$$

ima $\mathscr{P}(\sigma_n)$ distribuciju, gdje je σ_n definiran sa

$$\sigma_n = \sum_{j=1}^{n} \mu_j.$$

Dakle, za svaki $r \geq 0$ vrijedi

$$\mathbb{P}(S_n \leq r) = \sum_{k=1}^{n} \frac{\sigma_n^k \cdot e^{-\sigma_n}}{k!}.$$

Za fiksni r događaji $\{S_n \leq r\}$ čine opadajući niz kako n raste te njihov presjek iznosi $\{S \leq r\}$. Dakle,

$$\mathbb{P}(S \leq r) = \lim_{n\to\infty} \mathbb{P}(S_n \leq r) = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\sigma_n^k \cdot e^{-\sigma_n}}{k!}.$$

Ako σ_n konvergira prema nekom konačnom broju σ, tada neprekidnost funkcije $\pi_k(\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$ povlači da je

$$\mathbb{P}(S \leq r) = \sum_{k=0}^{r} \pi_r(\sigma)$$

te slijedi $\mathbb{P}(S = r) = \pi_r(\sigma)$. Zaključujemo da S ima $\mathscr{P}(\sigma)$ distribuciju. Ako $\sigma_n \to \infty$ slijedi

$$\mathbb{P}(S_n \leq r) = \sum_{k=1}^{n} \frac{\sigma_n^k \cdot e^{-\sigma_n}}{k!} = e^{-\sigma_n} \sum_{k=1}^{n} \frac{\sigma_n^k}{k!} \to 0,$$

iz čega slijedi da je $\mathbb{P}(S > r) = 1$, za $r \geq 0$ $\square$

Prethodni teorem pokazuje prirodnost proširenja koje smo uveli sa (1.1) i (1.2). Uz tu konvenciju slijedi da, ako slučajne varijable X_j imaju $\mathscr{P}(\mu_j)$ distribuciju, tada njihova suma ima $\mathscr{P}(\sum \mu_j)$ distribuciju neovisno da li ih imamo konačno ili prebrojivo mnogo te da li su neki od μ_j jednaki $0, \infty$ ili oboje.

Vektorski prostor S u kojem su smještene točke Poissonovog procesa su uobičajeno euklidski prostori dimenzije d. Međutim, teorija nas ne obavezuje da koristimo specijalna svojstva Euklidskih prostora, već samo trebamo familiju podskupova koristit kao testne skupove za brojanje slučajnih točaka, odnosno trebamo skupove za koje je *brojeća funkcija*

$$N(A) = \#\{\Pi \cap A\}$$

dobro definirana slučajna varijabla. Prirodno je pretpostaviti da je S izmjeriv prostor. Također, moramo osigurati da postoji dovoljno izmjerivih skupova za razlikovanje pojedinačnih točaka. To se može osigurati tako da pretpostavimo da je dijagonala

$$D = \{(x, y) \in S \times S : x = y\}$$

izmjeriva u prostoru $S \times S$. Rezultati teorije mjere nam potvrđuju valjanost navedene pretpostavke. To odmah upućuje da je svaki jednočlan skup $\{x\}$ u S izmjeriv. Navedene pretpostavke ćemo prešutno koristit u svim sljedećim tvrdnjama. Ukoliko je $S = \mathbb{R}^d$ tada uvijek uzimamo da su izmjerivi skupovi Borelovi (najmanja σ algebra koja sadrži sve otvorene skupove). Sada smo spremni definirati *Poissonov proces*.

Definicija 1.1.3. Poissonov proces *na S je proizvoljan prebrojiv podskup* Π *od S sa svojstvima:*

(i) Za proizvoljan skup međusobno disjunktnih podskupova $A_1, A_2, \ldots$ od S su slučajne varijable $N(A_1), N(A_2), \ldots$ nezavisne,

(ii) $N(A)$ ima Poissonovu distribuciju $\mathscr{P}(\lambda)$, gdje $\lambda = \lambda(A)$ leži u skupu $0 \leq \lambda \leq \infty$.

Dakle, ako je $\lambda(A)$ konačan broj, onda je i $\Pi \cap A$ također konačan skup s vjerojatnošću 1, odnosno prazan skup ako je $\lambda(A) = 0$. Ako je $\lambda(A) = \infty$, tada je $\Pi \cap A$ prebrojivo beskonačan sa vjerojatnošću 1. Također, ako su skupovi $A_1, A_2, \ldots$ međusobno disjunktni i u uniji daju A, tada je

$$N(A) = \sum_{n=1}^{\infty} N(A_n),$$

a očekivanje od $N(A)$ iznosi

$$\lambda(A) = \sum_{n=1}^{\infty} \lambda(A_n).$$

Slijedi, λ je mjera na S koju ćemo zvat *mjera očekivanja* (eng. *mean measure*) Poissonovog procesa Π. Mjera $N(\cdot)$ se zove *Poissonova slučajna mjera* (eng. *Poisson random measure*). Ako znamo mjeru očekivanja možemo, koristeći *(i) i (ii)*, odrediti sve zajedničke distribucije broja $N(A)$ za različite skupove A. Na primjer, uzmimo da su skupovi $A_1, A_2, \ldots$ izmjerivi, ali ne nužno disjunktni. Razmotrimo sve skupove oblika

$$B = A_1^* \cap A_2^* \cdots \cap A_n^*$$

gdje je svaki A_j^* jednak A_j ili njegovom komplementu. Skupova gornjeg oblika ima ukupno 2^n i međusobno su disjunktni. Ako ih napišemo sve kao $B_1, B_2, ..., B_{2^n}$, tada je

$$A_j = \bigcup_{i \in \gamma_j} B_i$$

gdje je γ_j podskup od $\{1, 2, ..., 2^n\}$. Stoga je $N(A) = \sum_{i \in \gamma_j} N(B_i)$ i $N(B_i)$ su nezavisni, sa Poissonovom razdiobom $\mathscr{P}(\lambda(B_i))$. Koristeći ova svojstva možemo izračunati vjerojatnost bilo kojeg događaja definiranog u terminima slučajnih varijabli $N(A_1), N(A_2), ..., N(A_n)$.

Ne može svaka mjera biti mjera očekivanja. Jedno od svojstava koje ćemo mi zahtijevati jest da je mjera jednočlanog skupa jednaka nuli, odnosno

$$\lambda(\{x\}) = 0.$$

Kada je $S = \mathbb{R}^d$ mjera očekivanja je najčešće dana u terminima gustoće. *Gustoća* je pozitivna, izmjeriva funkcija υ na S sa svojstvom da je

$$\lambda(A) = \int_A \upsilon(x) dx, \tag{1.4}$$

gdje dx označava $dx_1 dx_2 ... dx_d$. Ako je υ neprekidna funkcija u točki x, tada (1.4) povlači da za malu okolinu A od x možemo uzeti

$$\lambda(A) \sim \upsilon(x)|A|,$$

gdje $|A|$ označava *Lebesgueovu mjeru* skupa A. U posebnim slučajevima kada je λ konstanta, tj.

$$\lambda(A) = \upsilon \cdot |A|$$

govorimo o *uniformnom* ili *homogenom* Poissonovom procesu.

1.2 Teoremi o superpoziciji i transformaciji

Poissonov proces ima nekoliko posebnih svojstava koji čine izračun pripadajućih vjerojatnosti iznenađujuće jednostavnim. Dva najbitnija rezultata su svojstvo *superpozicije* i i svojstvo da je transformacija Poissonovog procesa ponovno Poissonov proces. Inače, svojstvo superpozicije je direktna posljedica teorema o prebrojivoj aditivnosti. Prije nego krenemo na dokaz svojstva superpozicije pokazat ćemo u sljedećoj lemi jedno (neobično) dodatno svojstvo.

Lema 1.2.1. *Neka su Π_1 i Π_2 dva nezavisna Poissonova procesa na skupu S, te neka je A izmjeriv skup s konačnim mjerama očekivanja $\mu_1(A)$ i $\mu_2(A)$. Tada su Π_1 i Π_2 međusobno disjunktni na skupu A s vjerojatnošću 1:*

$$\mathbb{P}(\Pi_1 \cap \Pi_2 \cap A = \emptyset) = 1. \tag{1.5}$$

Dokaz navedene leme se može pronaći u [2]. Rezultat se dalje proširuje na skupove A koji su prebrojiva unija skupova na kojima su $\mu_1(A)$ i $\mu_2(A)$ konačne mjere. Posebno, tvrdnja vrijedi i u slučaju da je $A = S$, uz uvjet da su $\mu_1(A)$ i $\mu_2(A)$ σ-konačne mjere.

Teorem 1.2.2. *(Teorem o superpoziciji)*
Neka $\Pi_1, \Pi_2, ...$ čine prebrojivi niz nezavisnih Poissonovih procesa na skupu S te neka je μ_n mjera očekivanja procesa Π_n, za svaki $n \in \mathbb{N}$. Tada je prebrojiva unija

$$\Pi = \bigcup_{n=1}^{\infty} \Pi_n$$

također Poissonov proces sa mjerom očekivanja

$$\mu = \sum_{n=1}^{\infty} \mu_n. \tag{1.6}$$

Dokaz. Neka je $N_n(A)$ broj točaka od Π_n koje se nalaze u izmjerivom skupu A. Ako je $\mu_n(A) < \infty$ za svaki $n \in \mathbb{N}$, tada lema 1.2.1 kaže da su skupovi Π_n disjunktni sa A te je broj točaka od Π u A dan sa

$$N(A) = \sum_{n=1}^{\infty} N_n(A). \tag{1.7}$$

Prema *teoremu o prebrojivoj aditivnosti* $N(A)$ ima $\mathscr{P}(\mu)$ razdiobu, gdje je mjera μ dana sa (1.6). S druge strane, ako je $\mu_n(A) = \infty$ za neki $n \in \mathbb{N}$, tada je $N_n(A) = N(A) = \infty$ i (1.7) je trivijalno ispunjeno. Kako bismo dokazali teorem dovoljno je pokazati da su $N(A_1), N(A_2), ..., N(A_k)$ nezavisne, pod uvjetom da su A_j međusobno disjunktni. To je jasno, jer su varijable

$$N_n(A_j),\ n = 1, 2, \dots,\ j = 1, 2, \dots, k \tag{1.8}$$

međusobno nezavisne, a $N(A_j)$ je definiran pomoću tih istih varijabli. Taj argument kompletira dokaz koje je, zbog općenitosti, iskazan u varijanti sa prebrojivim skupovima. Kao očiti korolar slijedi i tvrdnja za *konačne* skupove. □

Drugo važno svojstvo Poissonovog procesa jest da, ako se dani prostor preslika u drugi prostor, tako dobivene slučajne veličine opet formiraju Poissonov proces. Neka je Π Poissonov proces na danom prostoru S, μ mjera očekivanja te neka je $f : S \to T$. Pretpostavljamo da je T, kao i S, izmjeriv prostor i da je f izmjeriva funkcija u smislu da je

$$f^{-1}(B) = \{x \in S : f(x) \in B\}$$

izmjeriv podskup od S, za svaki izmjeriv podskup $B \subseteq T$. Točke $f(x)$, $x \in \Pi$ formiraju slučajan prebrojiv podskup $f(\Pi) \subseteq T$. Mi želimo pokazati da je to Poissonov proces. Označimo sa

$$N^*(B) = \#\{f(\Pi) \cap B\}$$

broj točaka od $f(\Pi)$ u skupu B. Dokle god su točke $f(x), x \in \Pi$ različite, slijedi da je

$$N^*(B) = \#\{x \in \Pi; f(x) \in B\} = N(f^{-1}(B)) \tag{1.9}$$

te (1.9) ima $\mathscr{P}(\mu^*)$ distribuciju, gdje je

$$\mu^* = \mu^*(B) = \mu(f^{-1}(B)) \tag{1.10}$$

Štoviše, ako su $B_1, B_2, ..., B_k \subseteq T$ međusobno disjunktni, tada su i njihove inverzne slike također međusobno disjunktni skupovi te, prema tome, slijedi da su slučajne varijable $N^*(B_j)$ nezavisne. Mjera očekivanja je dana sa (1.10) te je μ^* mjera inducirana iz μ sa funkcijom f. Međutim, uvjet da su točke $f(x)$ različite nije netrivijalan, jer ukoliko uzmemo da je f konstantna funkcija slijedi da će preslikati cijeli skup S, te implicitno i cijeli skup Π, u jednu točku u skupu T. Općenito, ako su μ i f takvi da inducirana mjera ima atom u točki $t \in T$, onda $A = f^{-1}\{t\}$ zadovoljava $m = \mu(A) = \mu^*(\{t\}) > 0$ te postoje barem dvije točke u Π koje padaju u A i koje se obje preslikavaju u točku t. Dakle, trebamo pretpostaviti da je μ^* bez atoma. Također nam je potreban uvjet o konačnosti za μ. Međutim, ispostavi se da je dovoljno da μ bude σ konačna, u smislu da se S može napisati kao prebrojiva unija izmjerivih skupova S_n koji su konačne mjere tj. $\mu(S_n) < \infty$

Teorem 1.2.3. *(Teorem o transformaciji)*
Neka je Π Poissonov proces sa σ - konačnom mjerom očekivanja na danom prostoru S, te neka je $f : S \to T$ izmjeriva funkcija takva da mjera inducirana sa f nema atoma. Tada je $f(\Pi)$ Poissonov proces na T s induciranom mjerom μ^* kao mjerom očekivanja.

Dokaz. Pretpostavimo najprije da je $\mu(S) < \infty$. Neka je $A \subseteq S$ proizvoljan izmjeriv skup te neka je A^C njegov komplement. Označimo sa Π_1, Π_2 restrikcije od Π obzirom na skup A, odnosno A^C, tako da Π_1, Π_2 predstavljaju nezavisne Poissonove procese. Budući su Π_1 i Π_2 nezavisni Poissonovi procesi, prema lemi 1.2.1, za $A = S$, slijedi da su Π_1 i Π_2 disjunktni sa vjerojatnošću 1. Na sličan način može se vidjeti (v. [2],odjeljak 2.3) da su skupovi $f(\Pi_1)$ i $f(\Pi_2)$ disjunktni sa vjerojatnošću 1. Označimo sa M mjeru na $S \times S$, gdje $M(C)$, $C \subseteq S \times S$ označava očekivani broj parova $(x, y) \in C$ za koje vrijedi

$$x \in \Pi,\ y \in \Pi,\ f(x) = f(y).$$

Tada nezavisnost od $f(\Pi_1)$ i $f(\Pi_2)$ povlači da je

$$M(A \times A^C) = 0$$

za svaki skup A. Svaki od skupova A i B možemo napisati kao disjunktnu unija dvaju drugih skupova na sljedeći način:

$$\begin{aligned} A &= (A \cap B) \cup (A \backslash B), \\ B &= (A \cap B) \cup (B \backslash A). \end{aligned}$$

Uvrštavanjem te korištenjem formule

$$(U \cup V) \times (U' \cup V') = (U \times U') \cup (U \times V') \cup (V \times U') \cup (V \times V')$$

slijedi:

$$\begin{aligned} M(A \times B) &= M\left(((A \cap B) \cup (A \backslash B)) \times ((A \cap B) \cup (B \backslash A))\right) \\ &= M\left(((A \cap B) \times (A \cap B)) \cup ((A \cap B) \times (B \backslash A)) \cup ((A \backslash B) \times (A \cap B)) \cup ((A \backslash B) \times (B \backslash A))\right) \\ &= M\left((A \cap B) \times (A \cap B)\right) + M\left((A \cap B) \times (B \backslash A)\right) + M\left((A \backslash B) \times (A \cap B)\right) + M\left((A \backslash B) \times (B \backslash A)\right) \\ &\leq M\left((A \cap B) \times (A \cap B)\right) + M\left(A \times A^C\right) + M\left(B^C \times B\right) + M\left(B^C \times B\right) \\ &= M\left((A \cap B) \times (A \cap B)\right) = m(A \cap B) \end{aligned}$$

gdje je $m(A) = M(A \times A)$. Zaključujemo kako je

$$M(A \times B) \leq m(A \cap B). \tag{1.11}$$

U prethodnom računu su korištene sljedeće relacije među skupovima:

$$\begin{aligned} (A \cap B) \times (B \backslash A) &\subseteq A \times A^C, \\ (A \backslash B) \times (A \cap B) &\subseteq B^C \times B, \\ (A \backslash B) \times (B \backslash A) &\subseteq B^C \times B. \end{aligned}$$

Budući da vrijedi

$$(A \cap B) \times (A \cap B) \subseteq A \times B$$

zaključujemo da je

$$m(A \cap B) \leq M(A \times B). \tag{1.12}$$

Iz (1.11) i (1.12) slijedi da je

$$m(A \cap B) = M(A \times B). \tag{1.13}$$

Stavljanjem $B = S$ u (1.13) dobivamo $m(A) = M(A \times S)$ iz čega slijedi da je m *mjera*. Neka je M_1 mjera inducirana pomoću mjere m na skupu $D = \{(x, x) : x \in S\} \subseteq S \times S$ tako da je

$$M_1(A \times B) = m(A \cap B) = M(A \times B),$$

za svaki A, B. Općenito, mjera je konačno aditivna funkcija. Budući je S izmjeriv skup zatvoren je na skupovne operacije. Stoga, prema rezultatima teorije mjere, postoji jedinstveno konačno aditivno proširenje pa zaključujemo kako je $M_1 = M$. Dakle, sa vjerojatnošću 1 zaključujemo da ne postoje različiti $x, y \in \Pi$ sa svojstvom da je $f(x) = f(y)$. To povlači (1.9) odakle slijedi tvrdnja teorema.

Napustimo sada pretpostavku da je $\mu(S) < \infty$ i pretpostavimo umjesto toga da postoje disjunktni skupovi S_n takvi da je

$$S = \bigcup_{n=1}^{\infty} S_n$$

uz uvjet da je $\mu(S_n) < \infty$. Označimo sa Π_n restrikciju od Π na S_n. Tada su Π_n nezavisni Poissonovi procesi sa konačnim mjerama očekivanja μ_n i $f(\Pi_n)$ su također nezavisni Poissonovi procesi sa konačnim mjerama očekivanja $\mu_n^*(B) = \mu_n(f^{-1}(B))$. Superpozicija

$$f(\Pi) = f\left(\bigcup_{n=1}^{\infty} \Pi_n\right) = \bigcup_{n=1}^{\infty} f(\Pi_n)$$

definira Poissonov proces sa mjerom očekivanja

$$\mu^*(B) = \sum_{n=1}^{\infty} \mu_n^*(B) = \mu(f^{-1}(B)).$$

Time je teorem dokazan. □

Kao ilustraciju primjene teorema pretpostavimo da je Π Poissonov proces na prostoru $\mathbb{R}^D$ s funkcijom gustoće $\lambda(x_1, x_2, ...x_D)$. Uzmimo da je $f : \mathbb{R}^D \to \mathbb{R}^d$ projekcija

$$f(x_1, x_2, ...x_D) = (x_1, x_2, ...x_d),$$

gdje je $d \leq D$. Za $B \subseteq \mathbb{R}^d$ imamo

$$\begin{aligned}\mu^*(B) &= \int \cdots \int_{B\times\mathbb{R}^{D-d}} \lambda(x_1, x_2, ...x_D)dx_1 \ldots dx_D \\ &= \int \cdots \int_{B} \lambda^*(x_1, x_2, ...x_d)dx_1 \ldots dx_d\end{aligned}$$

gdje je

$$\lambda^*(x_1, x_2, ...x_d) = \int \cdots \int_{\mathbb{R}^{D-d}} \lambda(x_1, x_2, ...x_D)dx_{d+1}dx_{d+2} \ldots dx_D. \tag{1.14}$$

Očito je μ σ-konačna i bez atoma tako da možemo primjeniti teorem te zaključujemo da, ako integral (1.14) konvergira, tada $f(\Pi)$ predstavlja Poissonov proces sa gustoćom λ^* na $\mathbb{R}^d$.

Sljedeće što nas zanima jest da li je moguće konstruirati Poissonov proces ukoliko imamo zadanu mjeru očekivanja. Svakako ćemo zahtijevati da je dana mjera bez atoma i da je konačna. Ispostavit će se da je uvjet o σ konačnosti dovoljan, ali ne i nužan.

Teorem 1.2.4. *(Egzistencija)*
Neka je μ mjera bez atoma na skupu S koja se može izraziti u obliku

$$\mu = \sum_{n=1}^{\infty} \mu_n, \quad \mu_n(S) < \infty.$$

Tada postoji Poissonov proces na skupu S takav da mu je μ mjera očekivanja.

Dokaz. Bez smanjenja općenitosti možemo pretpostaviti da je $\mu_n(S) > 0$ za svaki n. Na pogodnom vjerojatnosnom prostoru konstruiramo niz nezavisnih slučajnih varijabli

$$N_n, X_{nr} \quad (n, r = 1, 2, 3, ...)$$

tako da je distribucija od N_n jednaka $\mathscr{P}(\mu_n(S))$, a distribucija od X_{nr} je jednaka

$$p_n(\cdot) = \mu_n(\cdot)/\mu_n(S).$$

Označimo sa

$$\Pi_n = \{X_{n1}, X_{n2}, \ldots, X_{nN_n}\}$$

te sa
$$\Pi = \bigcup_{n=1}^{\infty} \Pi_n. \tag{1.15}$$
Označimo sa $N_n(A)$ broj elemenata skupa $\Pi_n \cap A$. Tada, za disjunktne $A_1, A_2, ..., A_k$ te za $A_0 = \left(\bigcup_{i=1}^{k} A_i\right)^C$ imamo
$$\begin{aligned}\mathbb{P}\{N_n(A_1) = m_1, N_n(A_2) = m_2, .., N_n(A_k) = m_k \mid N_n = m\} = \\ = \frac{m!}{m_0! m_1! \dots m_k!} p_n(A_0)^{m_0} p_n(A_1)^{m_1} \dots p_n(A_k)^{m_k},\end{aligned}$$
uz uvjet da je $m_0 = m - m_1 - m_2 - \cdots - m_k$. Slijedi
$$\begin{aligned}&\mathbb{P}\{N_n(A_1) = m_1, N_n(A_2) = m_2, .., N_n(A_k) = m_k\} = \\ &= \sum_{m=\sum m_j}^{\infty} \frac{e^{-\mu_n(S)} \mu_n(S)^m}{m!} \frac{m!}{m_0! m_1! \dots m_k!} \prod_{j=0}^{k} p_n(A_j^{m_j}) \\ &= \sum_{m=\sum m_j}^{\infty} \pi_{m-\sum m_j}(\mu_n(A_0)) \prod_{j=0}^{k} \pi_{m_j}(\mu_n(A_j)) \\ &= \prod_{j=0}^{k} \pi_{m_j}(\mu_n(A_j)).\end{aligned}$$
Prema tome, $N_n(A_j)$ su nezavisne slučajne varijable sa distribucijama $\mathscr{P}(\mu_n(A_j))$, a Π_n nezavisni Poissonovi procesi sa odgovarajućim mjerama očekivanja μ_n. Teorem o superpoziciji pokazuje da (1.15) definira Poissonov proces sa mjerom očekivanja μ čime je dokaz gotov. $\square$

1.3 Campbellov teorem

Pogledajmo pobliže sume oblike

$$\Sigma_f = \sum_{x \in \Pi} f(X)$$

gdje je f realna funkcija na danom vjerojatnosnom prostoru od Poissonovog procesa Π. Postoji značajan broj jednostavnih, ali važnih rezultata koji daju uvjete za (apsolutnu) konvergenciju takvih suma. Tehnika za pronalaženje distribucije od Σ_f jest da se ista prvo pronađe za jednostavne funkcije f te se potom rezultat proširi i u slučaju složenijih funkcija koristeći standardni postupak proširivanja.

Promotrimo prvo funkciju $f : S \to \mathbb{R}$ koja poprima konačno mnogo ne-nul vrijednosti $f_1, f_2, \ldots, f_k$ te je jednaka nuli izvan skupa na kojem je mjera očekivanja μ od Π konačna. Tada je

$$A_j = \{x; f(x) = f_j\},\ j = 1, \ldots, k$$

izmjeriv skup sa mjerom

$$m_j = \mu(A_j) < \infty$$

i A_j su međusobno disjunktni. Nadalje,

$$N_j = N(A_j)$$

su nezavisne slučajne varijable sa odgovarajućim Poissonovim distribucijama $\mathscr{P}(m_j)$ i

$$\Sigma_f = \sum_{X \in \Pi} f(X) = \sum_{j=1}^{k} f_j N_j. \tag{1.16}$$

Za proizvoljni realni ili kompleksni θ slijedi

$$\begin{aligned}
\mathbb{E}(e^{\theta \Sigma_f}) &= \prod_{j=1}^{k} \mathbb{E}(e^{\theta f_j N_j}) \\
&= \prod_{j=1}^{k} \exp\{(e^{\theta f_j} - 1) m_j\} \\
&= \exp\left\{ \sum_{j=1}^{k} \int_{A_j} (e^{\theta f(x)} - 1) \mu(dx) \right\} \\
&= \exp\left\{ \int_S (e^{\theta f(x)} - 1) \mu(dx) \right\}
\end{aligned}$$

jer je $f = 0$ izvan $\bigcup_j A_j$. Navedena jednakost

$$\mathbb{E}(e^{\theta \Sigma_f}) = \exp\left\{\int_S (e^{\theta f(x)} - 1)\mu(dx))\right\} \tag{1.17}$$

je temeljna jednakost u nastavku. Označimo ovu jednakost sa

$$g(\theta) = \exp\left\{\int_S (e^{\theta f(x)} - 1)\mu(dx)\right\}.$$

Tada su derivacije dane sa

$$\begin{aligned}
g'(\theta) &= \exp\left\{\int_S (e^{\theta f(x)} - 1)\mu(dx)\right\} \cdot \frac{d}{d\theta}\int_S (e^{\theta f(x)} - 1)\mu(dx) \\
&= g(\theta) \cdot \int_S \frac{d}{d\theta}(e^{\theta f(x)} - 1)\mu(dx) \\
&= g(\theta) \int_S \left(e^{\theta f(x)} f(x)\right)\mu(dx), && (1.18) \\
g''(\theta) &= g'(\theta) \cdot \int_S \left(e^{\theta f(x)} f(x)\right)\mu(dx) + g(\theta) \cdot \frac{d}{d\theta}\int_S \left(e^{\theta f(x)} f(x)\right)\mu(dx) \\
&= g(\theta)\left(\int_S \left(e^{\theta f(x)} f(x)\right)\mu(dx)\right)^2 + g(\theta)\int_S e^{\theta f(x)} f^2(x)\mu(dx). && (1.19)
\end{aligned}$$

Uvrštavanjem $\theta = 0$ u definiciju funkcije g te u jednakosti (1.18) i (1.19) dobivamo

$$\begin{aligned}
g(0) &= 1, \\
g'(0) &= g(0)\int_S f(x)\mu(dx) = \int_S f(x)\mu(dx), \\
g''(0) &= g(0)\left(\int_S f(x)\mu(dx)\right)^2 + g(0)\int_S f^2(x)\mu(dx) = \left(\int_S f(x)\mu(dx)\right)^2 + \int_S f^2(x)\mu(dx).
\end{aligned}$$

Razvojem u red funkcije $g(\theta)$ korištenjem formule

$$g(\theta) = \sum_{k=0}^{\infty} \frac{g^{(k)}(c)}{k!}(\theta - c)^k$$

te uvrštavanjem $c = 0$ dobivamo

$$g(\theta) = 1 + g'(0)\theta + \frac{g''(0)}{2!}\theta^2 + O_1(\theta).$$

Sa $O_1(\theta)$ smo označili preostale članove reda. Uvrštavanjem odgovarajućih vrijednosti dobivamo

$$g(\theta) = 1 + \theta \cdot \int_S f(x)\mu(dx) + \frac{\theta^2}{2!} \cdot \left(\left(\int_S f(x)\mu(dx) \right)^2 + \int_S f^2(x)\mu(dx) \right) + O_1(\theta). \quad (1.20)$$

Nadalje, razvijemo li lijevu stranu funkcije izvodnice momenata (1.17) u Taylorov red oko nule dobivamo

$$\begin{aligned} \mathbb{E}(e^{\theta\Sigma_f}) &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{\left(\theta\Sigma_f\right)^k}{k!} \right] \\ &= \sum_{k=0}^{\infty} \frac{\theta^k}{k!} \mathbb{E}\left(\Sigma_f^k\right) \\ &= 1 + \theta\mathbb{E}\left(\Sigma_f\right) + \frac{\theta^2}{2!}\mathbb{E}\left(\Sigma_f^2\right) + O_2(\theta), \end{aligned}$$

gdje smo sa $O_2(\theta)$ označili preostale članove reda. Izjednačavanjem koeficijenata uz θ i θ^2 dobivamo sljedeće identitete:

$$\begin{aligned} \mathbb{E}(\Sigma_f) &= \int_S f(x)\mu(dx), \\ \mathbb{E}\left(\Sigma_f^2\right) &= \left(\int_S f(x)\mu(dx) \right)^2 + \int_S f^2(x)\mu(dx). \end{aligned}$$

Uvrštavanjem u formulu za varijancu dobivamo da je

$$\mathrm{var}(\Sigma_f) = \int_S f(x)^2\mu(dx).$$

Ukoliko su f_1 i f_2 realne funkcije, zamjenom $\theta \mapsto \theta_1 f_1 + \theta_2 f_2$ u (1.17) dobivamo jednakost

$$\mathbb{E}(\mathrm{e}^{\theta_1\Sigma_1+\theta_2\Sigma_2}) = \exp\left\{ \int_S (e^{\theta_1 f_1(x)+\theta_2 f_2(x)} - 1)\mu(dx) \right\}, \quad (1.21)$$

koja određuje zajedničku distribuciju od

$$\Sigma_1 = \sum_{X\in\Pi} f_1(X), \quad \Sigma_2 = \sum_{X\in\Pi} f_2(X).$$

Rezultat se analogno može proširiti i na više od dvije funkcije. Razvijemo li (1.21) u red kao u prethodnom primjeru, tada će nam koeficijent uz $\theta_1\theta_2$ dati

$$\mathbb{E}(\Sigma_1\Sigma_2) = \int_S f_1 d\mu \int_S f_2 d\mu + \int_S f_1 f_2 d\mu. \quad (1.22)$$

Znamo da je

$$\mathbb{E}(\Sigma_1) = \int_S f_1 d\mu, \qquad \mathbb{E}(\Sigma_2) = \int_S f_2 d\mu.$$

Stoga drugi član u (1.22) predstavlja Cov (Σ_1, Σ_2). Ponekad je korisno napisati lijevu stranu kao

$$\mathbb{E}\left(\sum_{X_1, X_2 \in \Pi} f_1(X_1) f_2(X_2)\right) = \mathbb{E}\left(\sum_{X \in \Pi} f_1(X) f_2(X)\right) + \mathbb{E}\left(\sum_{X_1 \neq X_2} f_1(X_1) f_2(X_2)\right).$$

Prvi izraz na desnoj strani je $\int f_1 f_2 d\mu$ tako da iz (1.22) slijedi:

$$\mathbb{E}\left(\sum_{\substack{X_1, X_2 \in \Pi \\ X_1 \neq X_2}} f_1(X_1) f_2(X_2)\right) = \mathbb{E}\left(\sum_{X_1 \in \Pi} f_1(X_1)\right) \mathbb{E}\left(\sum_{X_2 \in \Pi} f_2(X_2)\right).$$

Navedeni identitet se može poopćiti i na više dimenzija:

$$\mathbb{E}\left(\sum_{\substack{X_1, X_2, \ldots, X_n \in \Pi \\ X_j \text{ različiti}}} f_1(X_1) f_2(X_2) \ldots f_n(X_n)\right) = \prod_{j=1}^{n} \mathbb{E}\left(\sum_{X_j \in \Pi} f_j(X_j)\right). \tag{1.23}$$

Formalni račun u prethodnom djelu je opravdan uzimanjem identiteta (1.17) za jednostavne funkcije f i proširivanjem u općenitije funkcije koristeći standardni postupak proširivanja poznat iz teorije mjere. Prethodni rezultati su skupljeni zajedno u idućem teoremu.

Teorem 1.3.1. *(Campbellov teorem)*
Neka je Π *Poissonov proces na skupu* S *sa mjerom očekivanja* μ *i neka je* $f : S \to \mathbb{R}$ *izmjeriva funkcija. Tada suma*

$$\Sigma_f = \sum_{X \in \Pi} f(X)$$

konvergira apsolutno sa vjerojatnošću 1 ako i samo ako vrijedi

$$\int_S \min(|f(x)|, 1) \mu(dx) < \infty. \tag{1.24}$$

Ako navedeni uvjet vrijedi, tada je

$$\mathbb{E}\left(e^{\theta \Sigma_f}\right) = \exp\left\{\int_S \left(e^{\theta f(x)} - 1\right) \mu(dx)\right\} \tag{1.25}$$

za proizvoljni $\theta \in \mathbb{C}$ za koji integral na desnoj strani jednakosti konvergira. Nadalje,

$$\mathbb{E}(\Sigma_f) = \int_S f(x)\mu(dx) \tag{1.26}$$

u smislu da očekivanje postoji ako i samo ako integral konvergira, a u tom slučaju vrijedi jednakost. Također, ako (1.26) *konvergira, tada je*

$$\operatorname{var}\left(\Sigma_f\right) = \int_S f(x)^2\mu(dx). \tag{1.27}$$

Dokaz. Znamo da (1.25) vrijedi za svaki $\theta \in \mathbb{C}$ ukoliko je f funkcija koja poprima konačan broj vrijednosti i nestaje izvan skupa sa konačnom μ-mjerom. Takve funkcije ćemo zvati *jednostavne funkcije*. Svaka pozitivna izmjeriva funkcija f se može izraziti u obliku rastućeg limesa jednostavnih funkcija (f_j). Uzimanjem $\theta = -u$ $(u > 0)$ te definiranjem $\Sigma_j = \sum f_j(X)$ imamo

$$\begin{aligned}
\mathbb{E}\left(e^{-u\Sigma_f}\right) &= \mathbb{E}\left(\lim_{j\to\infty} e^{-u\Sigma_j}\right)\\
&= \lim_{j\to\infty} \mathbb{E}\left(e^{-u\Sigma_j}\right)\\
&= \lim_{j\to\infty} \exp\left\{-\int_S \left(1 - e^{-uf_j(x)}\right)\mu(dx)\right\}\\
&= \exp\left\{-\int_S \left(1 - e^{-uf(x)}\right)\mu(dx)\right\}
\end{aligned}$$

gdje je zamjena limesa i integrala opravdana korištenjem Lebesgueovog teorema o monotonoj konvergenciji. Ako vrijedi (1.24),tada je

$$\int_S |f(x)|\mu(dx) < +\infty \qquad \text{ili} \qquad \mu(S) < +\infty.$$

Ako je $\int_S |f(x)|\mu(dx) < +\infty$, tada je

$$\left|\int_S f(x)\mu(dx)\right| < \int_S |f(x)|\mu(dx) < +\infty$$

iz čega slijedi da je $\int_S f(x)\mu(dx) < +\infty$ pa je $f(x) < +\infty$, za svaki $x \in S$. Stoga je

$$\lim_{u\to 0}\left(-\int_S \left(1 - e^{-uf(x)}\right)\mu(dx)\right) = 0$$

iz čega slijedi da je Σ_f konačna slučajna varijabla. Desnu stranu jednakosti (1.25) smo bili označili sa $g(\theta)$ i derivacija je dana sa (1.18). Sama funkcija $g(\theta)$ je neprekidna, za svaki $\theta = -u, u > 0$, i

$$\int_S \left(e^{\theta f(x)} f(x)\right) \mu(dx)$$

je također neprekidna za svaku konačnu funkciju f. Stoga su obje strane izraza (1.25) analitičke funkcije u ovisnosti o θ, za $\operatorname{Re}\theta \leq 0$, i (1.25) vrijedi za takav θ.

S druge strane, ukoliko (1.24) ne vrijedi, tada posljednji integral divergira za svaki $u > 0$ što povlači da je $\mathbb{E}\left(e^{-u\Sigma_f}\right) = 0$ te slijedi da je $\Sigma_f = \infty$ sa vjerojatnošću 1. Nadalje, znamo da za jednostavne funkcije vrijedi

$$\mathbb{E}\left(\Sigma_j\right) = \int_S f_j(x)\mu(dx).$$

Stoga uzimanjem limesa slijedi

$$\begin{aligned}\mathbb{E}\left(\Sigma_f\right) &= \mathbb{E}\left(\lim_{j\to\infty} \Sigma_j\right) \\ &= \lim_{j\to\infty} \mathbb{E}\left(\Sigma_j\right) \\ &= \lim_{j\to\infty} \int_S f_j \mu(dx) \\ &= \int_S f\mu(dx),\end{aligned}$$

iz čega slijedi jednakost (1.26). Analogna argumentacija vrijedi i za (1.27).

Dakle, teorem je dokazan ukoliko je $f \geq 0$. Kako bismo kompletirali dokaz potrebno je primjeniti teorem na pozitivne funkcije

$$f^+ = \max(f, 0) \quad , \quad f^- = \max(-f, 0).$$

Suma Σ_f konvergira apsolutno ako i samo ako sume

$$\begin{aligned}\Sigma_+ &= \sum_{X\in\Pi} f^+(X) = \sum_{X\in\Pi_+} f(X), \\ \Sigma_- &= \sum_{X\in\Pi} f^-(X) = -\sum_{X\in\Pi_-} f(X)\end{aligned}$$

konvergiraju, pri tom sa Π_+, Π_- označimo restrikcije od Π na skupove

$$S_+ = \{f > 0\}, \quad S_- = \{f < 0\}$$

respektivno. Time smo pokazali da je (1.24) nužan i dovoljan uvjet za konvergenciju. Ukoliko taj uvjet vrijedi i $\mathrm{Re}\,\theta = 0$, tada je

$$\begin{aligned}\mathbb{E}\left(e^{\theta\Sigma_f}\right) &= \mathbb{E}\left(e^{\theta\Sigma_+ - \theta\Sigma_-}\right)\\ &= \mathbb{E}\left(e^{\theta\Sigma_+}\right)\mathbb{E}\left(e^{-\theta\Sigma_-}\right)\\ &= \exp\left\{\int\left(e^{\theta f^+} - 1\right)d\mu\right\}\exp\left\{\int\left(e^{-\theta f^-} - 1\right)d\mu\right\}\\ &= \exp\left\{\int\left(e^{\theta f^+} + e^{-\theta f^-} - 2\right)d\mu\right\}\\ &= \exp\left\{\int\left(e^{\theta f} - 1\right)d\mu\right\},\end{aligned}$$

gdje smo iskoristili činjenicu da su Π_+, Π_- dva nezavisna Poissonova procesa koja u uniji daju cijeli Π. Dakle, (1.25) vrijedi za imaginarne θ. Analogno se rezultat proširi i na sve $\theta \in \mathbb{C}$. □

1.4 Subordinator

Započet ćemo poglavlje sa definicijom potpuno slučajne mjere koja je u uskoj vezi sa subordinatorom.

Definicija 1.4.1. *Kažemo da je slučajna mjera Φ potpuno slučajna mjera (eng. completely random measure) ako su, za svaku kolekciju međusobno nezavisnih skupova $A_1, A_2, \dots$ iz S, slučajne varijable $\Phi(A_1), \Phi(A_n), \dots$ međusobno nezavisne i*

$$\Phi\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \Phi(A_j).$$

Neka je Φ potpuno slučajna mjera na $\mathbb{R}$ koja je konačna na ograničenim skupovima. Tada takva mjera određuje slučajni proces $\phi : \mathbb{R} \to \mathbb{R}$ kroz preslikavanje

$$\phi(t) = \begin{cases}\Phi(0, t] & , t \geq 0\\ -\Phi(t, 0] & , t < 0.\end{cases}$$

Iz definicije je jasno da je ϕ rastuća i neprekidna zdesna te određuje slučajnu mjeru Φ na jedinstven način. Budući da je Φ nezavisna na konačnim skupovima slijedi da ϕ ima nezavisne priraste u smislu da su , za $t_1 < t_2 < \cdots < t_n$, prirasti

$$\phi(t_{j+1}) - \phi(t_j) \qquad (j = 1, \dots, n-1)$$

nezavisne slučajne varijable. Sada smo spremni uvesti definiciju subordinatora.

Definicija 1.4.2. *Proces* $\{\tau_s : s \geq 0\}$ *se zove subordinator ako ima nezavisne, stacionarne i nenegativne priraste* $\tau_t - \tau_s$, $s < t$ *pri čemu razdioba prirasta ovisi samo o razlici* $t - s$ *te vrijedi početni uvjet* $\tau_0 = 0$.

Homogen Poissonov proces $N(\cdot)$ na $\mathbb{R}$ je primjer slučajnog procesa koji je istovremeno potpuno slučajna mjera i subordinator. Naime, ako su $A_1, A_2, \dots$ disjunktni podskupovi od $\mathbb{R}$ tada, kao posljedica definicije, vrijedi

$$N\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=0}^{\infty} N(A_n),$$

što povlači da je $N(\cdot)$ potpuno slučajna mjera. S druge strane, budući je $N(0) = 0$, slijedi da je

$$N((j,t]) \sim \mathscr{P}(\lambda(t-j)),$$

a kako je $N((j,t]) = N((0,t]) - N((0,j]) = N_t - N_j$ zaključujemo da je

$$N_t - N_j \sim \mathscr{P}(\lambda(t-j)).$$

Stoga je $N(\cdot)$ subordinator. Jedan poznati subordinator koji će biti od koristi u izvodu Poisson - Dirichletove razdiobe jest *Moranov subordinator*. Općenito, to je subordinator čiji prirasti prate *gama* distribuciju. Kažemo da nenegativna slučajna varijabla Y ima *gama* distribuciju sa parametrima α i β, u oznaci $\mathscr{G}(\alpha,\beta)$, ako je funkcija gustoće dana sa

$$f(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}\,,\; y > 0.$$

Ukoliko je parametar $\beta = 1$, tada pišemo $\mathscr{G}(\alpha)$.

Definicija 1.4.3. *Moranov subordinator (ili Moranov proces)* $\left\{g_j : j \geq 0\right\}$ *je subordinator čiji prirasti* $g_s - g_t$, *za* $s < t$ *prate* $\mathscr{G}(s-t,1)$ *distribuciju.*

Definicija 1.4.4. *Kažemo da je stohastički proces* $\{X_t : t \geq 0\}$ *Lévyjev proces ako zadovoljava sljedeća svojstva:*

1. $X_0 = 0$ *gotovo sigurno,*
2. *za bilo koju subdiviziju* $0 \leq t_1 \leq \cdots \leq t_n < \infty$ *su prirasti* $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$ *nezavisni,*
3. *za bilo koji* $s < t$ *je radzioba od* $X_t - X_s$ *jednaka* X_{t-s},
4. *za bilo koji* $\epsilon > 0$ *i* $t \geq 0$ *vrijedi da je* $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| > \epsilon) = 0$.

Svaki Lévyjev proces je karakteriziran kroz svoju karakterističnu funkciju koja je dana Lévy - Hinčinovom formulom: ako je $X = (X_t)_{t\geq 0}$ Lévyjev proces, tada je karakteristična funkcija

$$\tau_X(\theta) = \mathbb{E}\left[e^{i\theta X_1}\right] = \exp\left\{ai\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}\setminus\{0\}}(e^{i\theta x} - 1 - i\theta x\mathbb{1}_{|x|<1})\Pi(dx)\right\},$$

gdje je $a \in \mathbb{R}, \sigma > 0, \mathbb{1}$ je karakteristična funkcija skupa $|x| < 1$ i Π je $\sigma-$ konačna mjera na $\mathbb{R}\setminus 0$ koja ima svojstvo da je $\int_{\mathbb{R}\setminus\{0\}} 1 \bigwedge x^2\Pi(dx) < \infty$. Uz Lévyjev proces se prirodno veže i pojam beskonačno djeljive vjerojatnosne distribucije. Kažemo da je vjerojatnosna distribucija beskonačno djeljiva (eng. *infinitely divisible*) ako se može izraziti kao vjerojatnosna distribucija sume proizvoljnog broja nezavisnih, jednakodistribuiranih slučajnih varijabli.

U našim daljnjim razmatranjima ćemo se koristit i pojmom Laplaceova transformacija čiju definiciju navodimo u nastavku.

Definicija 1.4.5. *Neka je $f : [0,\infty] \to \mathbb{C}$ funkcija. Laplaceova transformacija funkcije $f(t)$ je funkcija $F(s)$ definirana kao*

$$F(s) = \int_0^\infty e^{-st}f(t)dt, \qquad s \in \mathbb{C},$$

ukoliko taj integral postoji.

U teoriji vjerojatnosti se Laplaceova transformacija definira kao očekivana vrijednost tj. očekivanje: ako je X slučajna varijabla sa zadanom funkcijom gustoće f, tada je Laplaceova transformacija od f dana sa

$$F(s) = \mathbb{E}\left(e^{-sX}\right).$$

Poglavlje 2

Poisson - Dirichletova razdioba

U ovom odjeljku ćemo se usredotočit na središnju temu ovog rada - Poisson - Dirichletovu razdiobu. Uvest ćemo razdiobu i navesti primjer gdje se koristi. Prema Kingmanu, [2], distribucija se konstruira kroz gama proces.

2.1 Dirichletova distribucija

Često je potrebno promatrati slučajne vektore $p = (p_1, \dots, p_n)$ koji zadaju slučajnu diskretnu vjerojatnosnu distribuciju sa svojstvima

$$p_j \geq 0, \quad (j = 1, 2, \dots, n), \quad \sum_{j=1}^{n} p_j = n.$$

Na primjer, p_j može predstavljati proporciju neke j-te biološke populacije od njih ukupno n. Za svaki $K \geq 2$ definiramo K-dimenzionalni simpleks

$$\nabla_K = \left\{ (p_1, p_2, \dots, p_K) : p_i \geq 0, \sum_{j=1}^{K} p_j = 1 \right\}.$$

Uvjerljivo najjednostavniji netrivijalni primjer vjerojatnosne distribucije na simpleksu $\nabla_m, m \in \mathbb{N}$ jest *Dirichletova distribucija*, čiju definiciju navodimo u nastavku.

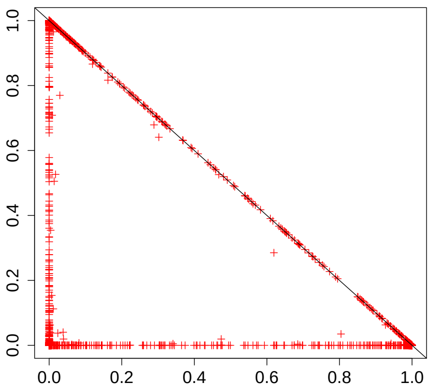
Definicija 2.1.1. *Za $m \geq 2$ Dirichletova distribucija, u oznaci $\mathscr{D}(\alpha_1, \dots, \alpha_m)$, je generalizacija beta distribucije sa više varijabli definirana na simpleksu ∇_m te ima sljedeću funkciju gustoće:*

$$f(p_1, \dots, p_{m-1}, p_m) = \frac{\Gamma\left(\sum_{k=1}^{m} \alpha_k\right)}{\prod_{k=1}^{m} \Gamma(\alpha_k)} p_1^{\alpha_1 - 1} \dots p_{m-1}^{\alpha_{m-1}-1} p_m^{\alpha_m - 1}, \tag{2.1}$$

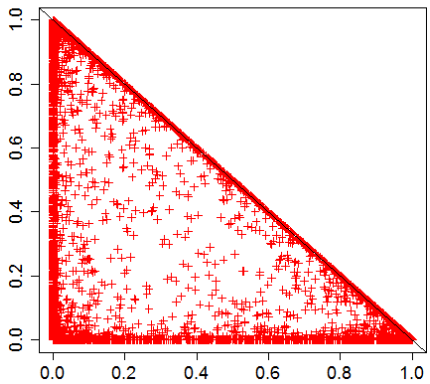
gdje su

$$(p_1, \dots, p_m) \in \nabla_m.$$

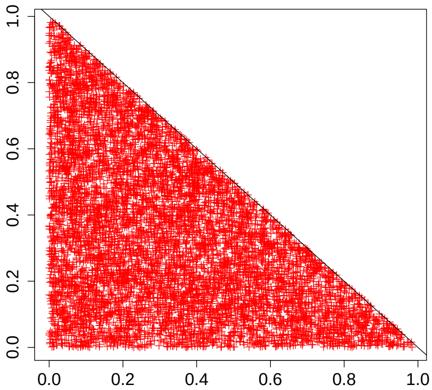
Parametri α_j mogu poprimiti bilo koju (strogo) pozitivnu vrijednost i karakter distribucije se značajno mijenja u ovisnosti o tim parametrima. Na primjer, ukoliko je $\alpha_j = 1$, za svaki $j = 1, \dots, m$, Dirichletova distribucija prelazi u uniformnu. Ukoliko su vrijednosti α_j velike, tada se dobivene vjerojatnosti uglavnom koncentriraju daleko od rubova simpleksa ∇_m. Sljedeće slike nam ilustriraju upravo navedena svojstva. U programskom paketu R su simulirani slučajni vektori $(p_1, p_2, p_3) \in \nabla_3$ sa Dirichletovom razdiobom za četiri različita izbora parametara. Mi smo grafički prikazali prve dvije komponente (p_1, p_2). Na apscisi je komponenta p_1, a na ordinati komponenta p_2.



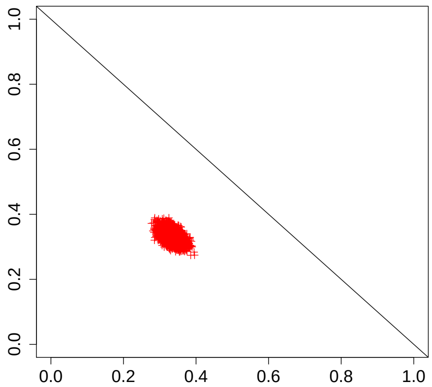
(a) $\mathscr{D}(0.01, 0.01, 0.01)$



(b) $\mathscr{D}(0.1, 0.1, 0.1)$



(c) $\mathscr{D}(1, 1, 1)$



(d) $\mathscr{D}(300, 300, 300)$

Slika 2.1: Simulirani slučajni vektori sa Dirichletovom razdiobom, duljina uzorka $n = 5000$.

U nastavku navodimo jednu relaciju između gama i Dirichletove distribucije koja će nam biti potrebna u našim budućim razmatranjima.

Lema 2.1.2. *Za svaki $m \geq 2$ označimo sa $Y_1, \dots, Y_m$ niz nezavisnih slučajnih varijabli sa $\mathcal{G}(\alpha_i)$, $i = 1, \dots, m$ razdiobom. Definirajmo slučajne varijable X_i , $i = 1, \dots, m$ na sljedeći način:*

$$X_i = \frac{Y_i}{\sum_{i=1}^m Y_k}.$$

Tada slučajni vektor $(X_1, \dots, X_m)$ ima $\mathcal{D}(\alpha_1, \dots, \alpha_m)$ distribuciju i nezavisan je sa $Y = \sum_{k=1}^m Y_k$.

Dokaz. Definirajmo sljedeću transformaciju vektora $(z, x_1, \dots, x_{m-1}) \in \mathbb{R}^m$:

$$(y_1, \dots, y_m) = T\,(z, x_1, \dots, x_{m-1}) = \left(zx_1, \dots, zx_{m-1}, z\left(1 - \sum_{i=1}^{m-1} x_i\right)\right).$$

Tada je Jakobijeva matrica $J(T)$ dana sa

$$J(T) = \begin{pmatrix} x_1 & z & 0 & \dots & 0 \\ x_2 & 0 & z & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m-1} & 0 & 0 & \dots & z \\ 1 - \sum_{i=1}^{m-1} x_i & -z & -z & \dots & -z \end{pmatrix}.$$

Kako bismo izračunali $\det(J(T))$ možemo pojednostavit matricu tako da prvih $m-1$ redaka dodamo zadnjem retku. Tako dobivena matrica ima jednaku determinantu kao i početna. Slijedi da je

$$\det(J(T)) = \det \begin{pmatrix} x_1 & z & 0 & \dots & 0 \\ x_2 & 0 & z & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m-1} & 0 & 0 & \dots & z \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix},$$

te Laplaceovim razvojem po zadnjem retku dobivamo da je $\det(J(T)) = z^{m-1}$. Nadalje, iz definicije funkcije T slijedi da je

$$x_i = \frac{y_i}{z}, \qquad i = 1, \dots, m-1.$$

Uzimanjem

$$z = \sum_{i=1}^m y_i$$

slijedi da je

$$\sum_{i=1}^{m} x_i = 1.$$

Stoga je zajednička gustoća $g(z, x_1, \dots, x_{m-1})$ od $(Y, X_1, \dots, X_{m-1})$ dana sa

$$\begin{aligned} g(z, x_1, \dots, x_{m-1}) &= \frac{1}{\Gamma(\alpha_1)\dots\Gamma(\alpha_m)} y_1^{\alpha_1 - 1} \dots y_m^{\alpha_m - 1} e^{-z} z^{m-1} \\ &= \frac{\Gamma(\sum_{k=1}^{m} \alpha_k)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_m)} x_1^{\alpha_1 - 1} \dots x_m^{\alpha_m - 1} \frac{1}{\Gamma(\sum_{k=1}^{m} \alpha_k)} z^{\sum_{k=1}^{m} \alpha_k - 1} e^{-z}, \end{aligned} \tag{2.2}$$

sto povlači da Y ima $\mathscr{G}(\sum_{k=1}^{m} \alpha_k, 1)$ distribuciju, $(X_1, \dots, X_m)$ ima $\mathscr{D}(\alpha_1, \dots, \alpha_m)$ distribuciju i nezavisan je sa Y. □

Izravno računanje razdiobom u (2.1) je složeno zbog zavisnosti između vrijednosti p_j. Poznato je da je lakše rukovati ovom distribucijom u terminima nezavisnih gama slučajnih varijabli. Prema lemi 2.1.2, slijedi da slučajni vektor $p = (p_1, \dots, p_n)$ s komponentama definiranim kao

$$p_j = \frac{Y_j}{Y} \tag{2.3}$$

ima $\mathscr{D}(\alpha_1, \dots, \alpha_n)$ razdiobu i nezavisan je sa Y. Također, Y ima $\mathscr{G}(\alpha)$ razdiobu, gdje je $\alpha = \alpha_1 + \dots + \alpha_n$. Navedeno također može bit dokazano koristeći Laplaceovu transformaciju

$$\mathbb{E}(e^{-\theta Y}) = \int_0^\infty g_\alpha(y) e^{-\theta y} dy = \frac{1}{(1+\theta)^\alpha}, \qquad \theta > -1,$$

gdje smo sa $g_\alpha(y)$ označili funkciju gustoće $\mathscr{G}(\alpha)$ razdiobe. Posebno, slijedi da je razdioba $\mathscr{G}(\alpha)$ beskonačno djeljiva te vrijedi jednakost

$$\frac{1}{(1+\theta)^\alpha} = \exp\left\{-\alpha \int_0^\infty (1 - e^{-\theta z}) z^{-1} e^{-z} dz\right\} \tag{2.4}$$

poznatija kao Lévy - Hinčin reprezentacija. Jednakosti (2.4) odgovara Moranov subordinator definiran u odjeljku 1.4.

Neka je $\left\{\phi_j : j \geq 0\right\}$ Moranov subordinator. Za $\alpha_1, \dots, \alpha_n > 0$ definiramo

$$t_0 = 0, \qquad t_j = \alpha_1 + \dots + \alpha_j, \qquad (1 \leq j \leq n).$$

Tada slučajne varijable $Y_j = \phi(t_j) - \phi(t_{j-1})$ imaju $\mathscr{G}(\alpha_j)$ distribuciju i Y_j su međusobno nezavisne. Budući da je

$$Y = Y_1 + \dots + Y_n = \phi(t_n) \sim \mathscr{G}(\alpha_1 + \dots + \alpha_n)$$

slijedi da smo sa

$$p_j = \frac{\phi(t_j) - \phi(t_{j-1})}{\phi(t_n)}$$

definirali slučajan vektor $p = (p_1, \dots, p_n) \in \nabla_n$ sa $\mathscr{D}(\alpha_1, \dots, \alpha_n)$ razdiobom.

2.2 Dirichletov proces

Moranov gama proces, uveden u odjeljku 1.4, odgovara potpuno slučajnoj mjeri Φ na skupu $\mathbb{R}$ sa svojstvom da $\Phi(A)$ ima $\mathscr{G}(\alpha)$ razdiobu, gdje je α Lebesgueova mjera skupa A. Neka je μ mjera na skupu S te promotrimo Poissonov proces Π^* na skupu $S^* = S \times (0, +\infty)$ čija mjera očekivanja je dana sa

$$\gamma(A, B) = \mu(A) \int_B z^{-1} e^{-z} dz,$$

gdje je $A \subseteq S, B \subseteq (0, +\infty]$. Za $A_1, A_2, \dots \subseteq S$ su slučajne varijable

$$\Phi(A_j) = \sum \left\{z; (x, z) \in \Pi^*, x \in A_j\right\}, \qquad j = 1, 2, \dots$$

međusobno nezavisne i vrijedi

$$\Phi\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \Phi(A_j).$$

Stoga je $\Phi(A), A \in S$ potpuno slučajna mjera čiju pripadajuću Laplaceovu transformaciju možemo izračunati koristeći Campbellov teorem i jednakost (2.4)

$$\begin{aligned}
\mathbb{E}(e^{-t\Phi(A)}) &= \exp\left\{-\int_{A^*} (1 - e^{-tz})\, \gamma(dz)\right\} \\
&= \exp\left\{-\int_0^{\infty} (1 - e^{-tz}) \mu(A) e^{-z} z^{-1} dz\right\} \\
&= \exp\left\{-\mu(A) \int_0^{+\infty} (1 - e^{-tz}) e^{-z} z^{-1} dz\right\} \\
&= \frac{1}{(1 + t)^{\mu(A)}}.
\end{aligned}$$

Dakle, $\Phi(A)$ ima $\mathscr{G}(\mu(A))$ razdiobu za svaki A za koji je $\mu(A)$ konačan. Neka su $A_1, \dots, A_n$ disjunktni skupovi, $A = \bigcup_{j=1}^{n} A_j$ i

$$\alpha_j = \mu(A_j) < \infty.$$

Tada su slučajne varijable

$$Y_j = \Phi(A_j)$$

nezavisne sa $\mathscr{G}(\alpha_j)$ distribucijom i

$$Y = Y_1 + \cdots + Y_n = \Phi(A).$$

Stoga (2.3) definira slučajni vektor $p \in \nabla_n$ čija distribucija je $\mathscr{D}(\alpha_1, \ldots, \alpha_n)$. Posebno, ukoliko je μ konačna mjera, tada je $\Phi(S)$ konačan sa vjerojatnošću 1 i

$$\Upsilon(A) = \Phi(A)/\Phi(S)$$

definira slučajnu vjerojatnosnu mjeru na skupu S takvu da je, za svaku particiju $A_1, \ldots, A_n$ od S, zajednička distribucija od

$$\Upsilon(A_1), \ldots, \Upsilon(A_n) \tag{2.5}$$

jednaka $\mathscr{D}(\alpha_1, \ldots, \alpha_n)$, uz $\alpha_j = \mu(A_j)$. Elementi u (2.5) nisu, dakako, nezavisni tako da Υ nije potpuno slučajna mjera. Mjeru Υ zovemo *Dirichletov proces koji odgovara mjeri* μ.

2.3 Izvod Poisson - Dirichletove razdiobe

Nakon digresije u općenite prostore vraćamo se u jednodimenzionalni kontekst. Pretpostavimo da slučajni vektor $p \in \nabla_n$ ima $\mathscr{D}(\alpha_1, \ldots, \alpha_n)$ razdiobu gdje, zbog određenosti, pretpostavljamo da su $\alpha_j = \alpha$, za sve $j = 1, \ldots, n$. Koristeći simetriju vidimo da je

$$\mathbb{E}(p_j) = \frac{1}{n},$$

tako da, kad je n velik, svi su p_j tipično male vrijednosti. To je istinito bez obzira na vrijednost α. Sa

$$p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(n)} \tag{2.6}$$

označimo vrijednosti p_j u padajućem poretku. Može li se išta reći o $p_{(1)}$ ili, općenitije, o $p_{(k)}$ i o slučajnom vektoru $(p_{(1)}, \ldots, p_{(k)})$ za fiksirani k uz $n \to \infty$ i $\alpha \to 0$? Odgovor je potvrdan, uz uvjet da $n\alpha$ konvergira prema nekoj konačnoj vrijednosti $\lambda \neq 0$.

Pretpostavimo, općenitije, da (2.6) predstavlja elemente slučajnog vektora $p^{(n)}$, u padajućem poretku, sa Dirichletovom distribucijom

$$\mathscr{D}(\alpha_1^{(n)}, \ldots, \alpha_n^{(n)}).$$

Pretpostavimo da, kako $n \to \infty$,

$$\max(\alpha_1^{(n)}, \dots, \alpha_n^{(n)}) \to 0 \tag{2.7}$$

i da

$$\lambda^{(n)} = \alpha_1^{(n)} + \cdots + \alpha_n^{(n)} \to \lambda. \tag{2.8}$$

Tada postoje granični rezultati za $p_{(k)}^{(n)}$ za fiksni k kako $n \to \infty$. Neka je $\phi(t)$ Moranov proces. Za $j = 1, \dots, n$ definirajmo

$$\bar{p}_j^{(n)} = \frac{\phi(\alpha_1^{(n)} + \cdots + \alpha_j^{(n)}) - \phi(\alpha_1^{(n)} + \cdots + \alpha_{j-1}^{(n)})}{\phi(\lambda^{(n)})}.$$

Tada, prema lemi 2.1.2, slučajni vektor $\bar{p}^{(n)} = (\bar{p}_1^{(n)}, \dots, \bar{p}_n^{(n)})$ ima istu distribuciju kao i $p^{(n)}$ i distribucija od $\bar{p}_{(k)}^{(n)}$ je ista kao i od $p_{(k)}^{(n)}$. No, uz uvjete (2.7) i (2.8), $\bar{p}_{(k)}^{(n)}$ konvergira po distribuciji, za $n \to \infty$. Nadalje, označimo sa $J_1 \geq J_2 \geq J_3 \geq \dots$ veličine skokova procesa ϕ na intervalu $[0, \lambda]$ u padajućem poretku. Tada vrijedi

$$\lim_{n\to\infty} \phi(\lambda^{(n)}) \bar{p}_{(k)}^{(n)} = J_k,$$

gdje konvergencija slijedi po distribuciji. Može se pokazati (v. [2]) da, za svaki k, zajednička distribucija od

$$\bar{p}_{(1)}^{(n)}, \dots, \bar{p}_{(k)}^{(n)},$$

a posljedično i od

$$p_{(1)}^{(n)}, \dots, p_{(k)}^{(n)},$$

konvergira prema distribuciji od $\xi_1, \xi_2, \dots, \xi_k$, gdje je

$$\xi_k = \frac{J_k}{\phi(\lambda)}.$$

Budući da je

$$\begin{aligned}
\phi(\lambda) &= \lim_{n\to\infty} \phi(\lambda^{(n)}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \Big(\phi(\alpha_1^{(n)} + \cdots + \alpha_k^{(n)}) - \phi(\alpha_1^{(n)} + \cdots + \alpha_{k-1}^{(n)})\Big) \\
&= \sum_{k=1}^{\infty} J_k,
\end{aligned}$$

slijedi

$$\xi_1 \geq \xi_2 \geq \dots, \quad \sum_{k=1}^{\infty} \xi_k = 1. \tag{2.9}$$

Distribucija beskonačnog niza slučajnih varijabli $\xi = (\xi_1, \xi_2, \dots)$ koja zadovoljava (2.9) ovisi samo o λ i zove se *Poisson - Dirichlet razdioba*, u oznaci $\mathscr{PD}(\lambda)$. Rezultat možemo sažeti tako da kažemo da uređajne statistike slučajnog vektora s $\mathscr{D}(\alpha_1, \dots, \alpha_n)$ razdiobom teži prema $\mathscr{PD}(\lambda)$ razdiobi sve dok je n velik, α_j jednoliko maleni i $\alpha_1 + \cdots + \alpha_n$ je približno jednaka λ. Ispostavlja se da je to od velike važnosti u mnogim primjenama, posebice u populacijskoj genetici i ekologiji. Dirichletova razdioba je ravnotežna razdioba za razne evolucijske modele i kombinacija malih α_j sa velikim n se često pojavljuje.

Sljedeći primjer ilustrira primjenu Poisson - Dirichletove razdiobe u teoriji brojeva. Detalji se mogu pronaći u [1].

Primjer 2.3.1. *(Rastav na proste faktore)*
Za svaki $n \in \mathbb{N}$ sa N_n označimo slučajno izabran prirodan broj od $1, 2, \dots, n$. Promotrimo rastav na proste faktore

$$N_n = \prod_p p^{C_p(n)}$$

gdje smo sa $C_p(n)$ označili kratnost od p. Definiramo

$$\begin{aligned} K_n &= \sum_p C_p(n), \\ A_i(n) &= i\text{-ti najveći prosti faktor od } N_n, \; i = 1, \dots, K_n \\ A_i(n) &= 1, \qquad i > K_n \\ L_i(n) &= \log A_i(n), \qquad i \geq 1 \end{aligned}$$

Kako $n \to \infty$, tako $\left(\frac{L_1(n)}{\log n}, \frac{L_2(n)}{\log n}, \dots\right)$ konvergira po distribuciji prema $(Y_1, Y_2, \dots)$ koji ima Poisson - Dirichletovu razdiobu sa parametrom 1.

Poisson - Dirichletova razdioba je razdioba beskonačnog niza $\xi = (\xi_1, \xi_2, \dots)$ koji zadovoljava (2.9). Takav niz, sa navedenom razdiobom, se može generirati kroz formulu

$$\xi_k = J_k(\lambda)/\phi(\lambda)$$

gdje $J_k(\lambda)$ označava k-ti najveći skok na segmentu $[0, \lambda]$ Moranovog procesa ϕ. Ispostavlja se da pozicije i veličine tih skokova formiraju Poissonov proces Π^* čija mjera očekivanja ima gustoću

$$\mu^*(dz) = z^{-1} e^{-z} dz, \qquad z > 0.$$

Nadalje, prema teoremu o transformaciji, veličine skokova $J_k(\lambda)$ na segmentu $[0, \lambda]$ formiraju Poissonov proces Π_λ na $(0, +\infty)$ čija mjera očekivanja iznosi

$$\lambda z^{-1} e^{-z}. \tag{2.10}$$

Uočimo, $J_k(\lambda)$ su točke od Π_λ u padajućem poretku i

$$\phi(\lambda) = \sum_{k=1}^{\infty} J_k(\lambda) < \infty.$$

Na ovaj način se Poisson - Dirichletova razdioba može opisati preko Poissonovog procesa Π_λ sa mjerom očekivanja (2.10). Ako su $\eta_1 \geq \eta_2 \geq \dots$ točke takvog procesa tada, prema Campbellovom teoremu, slijedi da je

$$\sigma = \sum_{k=1}^{\infty} \eta_k < \infty$$

i

$$\xi_k = \eta_k/\sigma \tag{2.11}$$

definira niz ξ koji ima $\mathscr{PD}(\lambda)$ razdiobu. Korisno je u nastavku primjetiti kako su ξ i λ međusobno nezavisni što proizlazi iz nezavisnosti vektora p i Y u odjeljku 2.2 tj. jednakosti (2.3). Također, budući da

$$\sigma_n = \sum_{k=1}^{n} \eta_k \sim \mathscr{G}(\lambda^{(n)})$$

slijedi da, puštanjem limesa kada $n \to \infty$,

$$\sigma = \sum_{k=1}^{\infty} \eta_k \sim \mathscr{G}(\lambda).$$

2.4 Marginalna distribucija

U ovom odjeljku ćemo nešto više reći o marginalnoj distribuciji Poisson - Dirichletove razdiobe. Uzmimo da slučajni vektor $(\bar{p}_1^{(n)}, \dots, \bar{p}_n^{(n)}$ ima $\mathscr{D}(\alpha, \dots, \alpha)$ razdiobu te neka padajuće uređen slučajni vektor $\xi = (\xi_1, \xi_2, \dots)$ ima $\mathscr{PD}(\lambda), \lambda > 0$ distribuciju. Za početak ćemo dokazati jednu pomoćnu lemu koja će biti od koristi u narednim razmatranjima.

Lema 2.4.1. *Neka slučajni vektor* $(p_1, \dots, p_n)$ *ima* $\mathscr{D}(\alpha_1, \dots, \alpha_n)$ *razdiobu te neka su* $A_1, \dots, A_r$ *disjunktne particije skupa* $\{1, 2, \dots, n\}$. *Tada vrijedi:*

$$\left(\sum_{i\in A_1} p_i, \dots, \sum_{i\in A_r} p_i\right) \sim \mathscr{D}\left(\sum_{i\in A_1} \alpha_i, \dots, \sum_{i\in A_r} \alpha_i\right).$$

Dokaz. U nastavku ćemo se oslanjati na svojstvo da, ako je $X_i \sim \mathscr{G}(\kappa_i, \delta)$ za $i = 1, \dots, n$, onda vrijedi

$$\sum_{i=1}^{n} X_i \sim \mathscr{G}\left(\sum_{i=1}^{n} \kappa_i, \delta\right).$$

Budući da slučajni vektor $(p_1, \dots, p_n)$ ima $\mathscr{D}(\alpha_1, \dots, \alpha_n)$ razdiobu, znamo da postoji niz nezavisnih slučajnih varijabli $Y_i \sim \mathscr{G}(\alpha_i, 1)$ tako da se svaki $p_i, i = 1, \dots, n$ može napisati kao

$$p_i = \frac{Y_i}{Y},$$

gdje je $Y = \sum_{i=1}^{n} Y_i$. Tada slijedi

$$\left(\sum_{i \in A_1} p_i, \dots, \sum_{i \in A_r} p_i\right) = \frac{1}{Y}\left(\sum_{i \in A_1} Y_i, \dots, \sum_{i \in A_r} Y_i\right).$$

Prema uvodnoj napomeni u dokazu znamo da

$$\begin{aligned}
\sum_{i \in A_1} Y_i &\sim \mathscr{G}\left(\sum_{i \in A_1} \alpha_i, 1\right), \\
&\vdots \\
\sum_{i \in A_r} Y_i &\sim \mathscr{G}\left(\sum_{i \in A_r} \alpha_i, 1\right), \\
Y &\sim \mathscr{G}\left(\sum_{i=1}^{n} \alpha_i, 1\right).
\end{aligned}$$

Tvrdnja leme sada slijedi iz leme 2.1.2. □

Kao korolar prethodne leme slijedi da, ako slučajni vektor$(p_1, \dots, p_n)$ ima $\mathscr{D}(\alpha_1, \dots, \alpha_n)$ razdiobu, tada

$$(p_1, \dots, p_j, p_{j+1} + \dots + p_n) \sim \mathscr{D}(\alpha_1, \dots, \alpha_j, \alpha_{j+1} + \dots + \alpha_n). \tag{2.12}$$

Neka je $\alpha = \lambda/n$. Sada smo spremni pokazati sljedeću propoziciju.

Propozicija 2.4.2. *Za $x > 0$ vrijedi*

$$\mathbb{P}(\xi_1 \le x) = 1 + \sum_{j=1}^{[1/x]} \frac{(-\lambda)^j}{j!} \int_x^1 \dots \int_x^1 \frac{(1 - y_1 - \dots - y_j)_+^{\lambda-1}}{y_1 y_2 \dots y_j} dy_1 \dots dy_j.$$

Dokaz. Za $0 < x < 1$, koristeći formulu uključivanja - isključivanja, simetriju te svojstva Dirichletove razdiobe, imamo da je

$$\begin{aligned}\mathbb{P}(\bar{p}_{(1)}^{(n)} \leq x) &= 1 - \mathbb{P}(\bar{p}_1^{(n)} > x \cup \cdots \cup \bar{p}_n^{(n)} > x) \\ &= 1 + \sum_{j=1}^{[1/x]} (-1)^j \binom{n}{j} \mathbb{P}(\bar{p}_1^{(n)} > x \cap \cdots \cap \bar{p}_j^{(n)} > x).\end{aligned}$$

Koristeći svojstvo (2.12), uz uvjet da su $\alpha_i = \alpha, i = 1, \ldots, n$ i $\alpha = \frac{\lambda}{n}$ slijedi da

$$\left(\bar{p}_1^{(n)}, \ldots, \bar{p}_j^{(n)}, \bar{p}_{j+1}^{(n)} + \cdots + \bar{p}_n^{(n)}\right) \sim \mathscr{D}\left(\alpha, \ldots, \alpha, (n-j)\alpha\right).$$

Primjetimo kako prethodni vektor ima ukupno $j + 1$ elemenata. Stoga, prema definiciji Dirichletove razdiobe, slijedi

$$\begin{aligned}&\mathbb{P}(\bar{p}_1^{(n)} > x \cap \cdots \cap \bar{p}_j^{(n)} > x) = \\ &= \int_x^1 \cdots \int_x^1 \frac{\Gamma(j\alpha + (n-j)\alpha)}{\Gamma(\alpha)^j \Gamma((n-j)\alpha)} y_1^{\alpha-1} \ldots y_j^{\alpha-1} (1 - y_1 - \cdots - y_j)^{(n-j)\alpha-1} dy_1 \ldots dy_j \\ &= \int_x^1 \cdots \int_x^1 \frac{\Gamma(n\alpha)(1 - y_1 - \cdots - y_j)^{n\alpha - j\alpha - 1}}{\Gamma(\alpha)^j \Gamma(n\alpha - j\alpha)} y_1^{\alpha-1} \ldots y_j^{\alpha-1} dy_1 \ldots dy_j.\end{aligned}$$

Sada slijedi

$$\mathbb{P}(\bar{p}_{(1)}^{(n)} \leq x) = \sum_{j=1}^{[1/x]} (-1)^j \binom{n}{j} \alpha^j \int_x^1 \cdots \int_x^1 \frac{\Gamma(n\alpha)(1 - y_1 - \cdots - y_j)^{n\alpha - j\alpha - 1}}{\alpha^j \Gamma(\alpha)^j \Gamma(n\alpha - j\alpha)} y_1^{\alpha-1} \ldots y_j^{\alpha-1} dy_1 \ldots dy_j.$$

Dakle, za $n \to \infty$ i $\alpha = \lambda/n \to 0$ imamo da je

$$\begin{aligned}\binom{n}{j}\alpha^j &= \frac{n(n-1)\ldots(n-j+1)}{j!}\alpha^j \\ &= \frac{\alpha n(\alpha n - \alpha)\ldots(\alpha n - \alpha j + \alpha)}{j!} \to \lambda^j / j!, \\ &\alpha\Gamma(\alpha) = \Gamma(\alpha + 1) \to 1,\end{aligned}$$

a kako $\mathbb{P}(\bar{p}_{(1)}^{(n)} \leq x) \to \mathbb{P}(\xi_1 \leq x)$, slijedi

$$\mathbb{P}(\bar{p}_{(1)}^{(n)} \leq x) \to 1 + \sum_{j=1}^{[1/x]} \frac{(-\lambda)^j}{j!} \int_x^1 \cdots \int_x^1 \frac{(1 - y_1 - \cdots - y_j)_+^{\lambda-1}}{y_1 y_2 \ldots y_j} dy_1 \ldots dy_j$$

čime je pokazana tvrdnja. $\square$

U nastavku navodimo zajedničku gustoću slučajnog vektora $(\xi_1, \dots, \xi_r)$.

Propozicija 2.4.3. *Zajednička gustoća slučajnog vektora $(\xi_1, \dots, \xi_r)$ je dana sa*

$$f_{\xi_1,\dots,\xi_r}(z_1, \dots, z_r) = \frac{\lambda^r(1 - z_1 - \dots - z_r)^{\lambda-1}}{z_1 \cdots z_r} \cdot \mathbb{P}\left(\xi_1 \leq \frac{z_r}{1 - z_1 - \dots - z_r}\right),$$

gdje su $z_1 > z_2 > \dots > z_r > 0$ i $z_1 + z_2 + \dots + z_r < 1$.

Dokaz. Promotrimo gustoću $f(x_1, \dots, x_r)$ prvih r uređenih statistika između $(Z_1, \dots, Z_n) \sim \mathscr{D}(\alpha, \dots, \alpha)$ u padajućem poretku. Tada postoji

$$n(n-1)\cdots(n-r+1)$$

mogućih varijacija za odabir. Slijedi

$$f(x_1, \dots, x_r) = n(n-1)\cdots(n-r+1) \int \dots \int g(x_1, \dots, x_n) dx_{r+1} \cdots dx_{n-1}$$

gdje funkcija g označava gustoću Dirichletove razdiobe. Prethodni integral drži varijable $x_1, \dots, x_r$ fiksnima, uz uvjet

$$1 \geq x_1 \geq x_2 \geq \dots \geq x_r \geq 0.$$

Varijable $x_1, \dots, x_r$ ćemo u nastavku označavati sa $z_1, \dots, z_r$. Neka su $Z_{(1)} > Z_{(2)} > \dots > Z_{(n)}$ uređajne statistike slučajnog vektora $Z = (Z_1, \dots, Z_n)$ sa $\mathscr{D}(\alpha, \dots, \alpha)$ razdiobom. Integrirajući po skupu

$$\begin{aligned} B &= \{(x_{r+1}, \dots, x_n) : 0 \leq x_i < z_r, z_1 + \dots + z_r + x_{r+1} + \dots + x_n = 1\} \\ &= \{(x_{r+1}, \dots, x_n) : 0 \leq x_i < z_r, x_{r+1} + \dots + x_n = 1 - z_1 - \dots - z_r\}, \end{aligned}$$

vidimo da se gustoća od $(Z_{(1)}, \dots, Z_{(r)})$ može napisati kao

$$\int_B \frac{n(n-1)\cdots(n-r+1)\Gamma(\lambda)}{\Gamma(\alpha)^n} z_1^{\alpha-1} \cdots z_r^{\alpha-1} x_{r+1}^{\alpha-1} \cdots x_n^{\alpha-1} dx_{r+1} \cdots dx_{n-1}. \tag{2.13}$$

Nadalje, vrijedi

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

iz čega slijedi jednakost

$$\Gamma(\alpha)^r = \frac{\Gamma(\alpha+1)^r}{\alpha^r}. \tag{2.14}$$

Uvrštavanjem (2.14) u jednakost (2.13) slijedi

$$\begin{aligned}
&\frac{n(n-1)\cdots(n-r+1)\Gamma(\lambda)}{\Gamma(\alpha)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\cdots z_r^{\alpha-1}\int_B \frac{\Gamma(\lambda-r\alpha)}{\Gamma(\alpha)^{n-r}} x_{r+1}^{\alpha-1}\cdots x_n^{\alpha-1} dx_{r+1}\cdots dx_{n-1}\\
&= \frac{n(n-1)\cdots(n-r+1)\Gamma(\lambda)\alpha^r}{\Gamma(\alpha+1)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\cdots z_r^{\alpha-1}\int_B \frac{\Gamma(\lambda-r\alpha)}{\Gamma(\alpha)^{n-r}} x_{r+1}^{\alpha-1}\cdots x_n^{\alpha-1} dx_{r+1}\cdots dx_{n-1}.
\end{aligned}$$

Zamjenom varijabli

$$\begin{aligned}
y_j &= \frac{x_j}{1-z_1-\cdots-z_r}, \qquad j = r+1,\ldots,n\\
dy_j &= \frac{dx_j}{1-z_1-\cdots-z_r},
\end{aligned}$$

gustoća postaje

$$\begin{aligned}
&\frac{n(n-1)\cdots(n-r+1)\alpha^r\Gamma(\lambda)}{\Gamma(\alpha+1)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\cdots z_r^{\alpha-1}\cdot\\
&\cdot\int_B \frac{\Gamma(\lambda-r\alpha)}{\Gamma(\alpha)^{n-r}} y_{r+1}^{\alpha-1}(1-z_1-\cdots-z_r)^{\alpha-1}\ldots y_n^{\alpha-1}(1-z_1-\cdots-z_r)^{\alpha-1}\\
&\cdot(1-z_1-\cdots-z_r)^{n-r-1} dy_{r+1}\ldots dy_{n-1}\\
&= \frac{n(n-1)\cdots(n-r+1)\alpha^r\Gamma(\lambda)}{\Gamma(\alpha+1)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\ldots z_r^{\alpha-1}\cdot\\
&\cdot\int_B \frac{\Gamma(\lambda-r\alpha)}{\Gamma(\alpha)^{n-r}} (y_{r+1}\cdots y_n)^{\alpha-1}(1-z_1-\cdots-z_r)^{(n-r)(\alpha-1)+n-r-1} dy_{r+1}\ldots dy_n\\
&= \frac{n(n-1)\cdots(n-r+1)\alpha^r\Gamma(\lambda)}{\Gamma(\alpha+1)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\ldots z_r^{\alpha-1}\cdot\\
&\cdot\int_B \frac{\Gamma(\lambda-r\alpha)}{\Gamma(\alpha)^{n-r}} (y_{r+1}\cdots y_n)^{\alpha-1}(1-z_1-\cdots-z_r)^{(n-r)\alpha-1} dy_{r+1}\cdots dy_n\\
&= \frac{n(n-1)\cdots(n-r+1)\alpha^r\Gamma(\lambda)}{\Gamma(\alpha+1)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\ldots z_r^{\alpha-1}(1-z_1-\cdots-z_r)^{(n-r)\alpha-1}\\
&\cdot\int_B \frac{\Gamma(\lambda-r\alpha)}{\Gamma(\alpha)^{n-r}} (y_{r+1}\ldots y_n)^{\alpha-1} dy_{r+1}\ldots dy_n\\
&= \frac{n(n-1)\cdots(n-r+1)\alpha^r\Gamma(\lambda)}{\Gamma(\alpha+1)^r\Gamma(\lambda-r\alpha)} z_1^{\alpha-1}\ldots z_r^{\alpha-1}(1-z_1-\cdots-z_r)^{(n-r)\alpha-1}\\
&\cdot\mathbb{P}\left(\max_{r+1\le j\le n} Z'_j \le \frac{z_r}{1-z_1-\cdots-z_r}\right),
\end{aligned}$$

gdje $\left(Z'_{r+1},\ldots,Z'_n\right) \sim \mathscr{D}(\alpha,\ldots,\alpha)$. Puštanjem limesa kada $n\to\infty$ slijedi

$$f_{Z_{(1)},\ldots,Z_{(r)}}(z_1,\ldots,z_r) \to \frac{\lambda^r(1-z_1-\cdots-z_r)^{\lambda-1}}{z_1\ldots z_r}\mathbb{P}\left(\xi_1 \le \frac{z_r}{1-z_1-\cdots-z_r}\right).$$

Budući da $(Z_{(1)}, \dots, Z_{(r)}) \to (\xi_1, \dots, \xi_r)$ slijedi tvrdnja. $\square$

2.5 Ewensova formula uzorkovanja

Zbog njene veze sa Dirichletovom razdiobom, Poisson - Dirichletova razdioba se često pojavljuje kao model za podjelu populacije sa velikim brojem članova prema vrsti ili nekom određenom svojstvu. Tada beskonačan slučajni vektor

$$\xi = (\xi_1, \xi_2, \dots) \sim \mathscr{P}\mathscr{D}(\lambda) \tag{2.15}$$

opisuje strukturu populacije, gdje ξ_k predstavlja proporciju k-tog tipa. Za ilustraciju pretpostavimo da smo iz takve velike populacije na slučajan način uzeli uzorak veličine n. Kolika je vjerojatnost da su svi elementi uzorka istog tipa? Ako uvjetujemo na ξ, vjerojatnost je jednaka

$$\sum_{k=1}^{\infty} \xi_k^n$$

iz čega slijedi da je vjerojatnost bez uvjeta na ξ dana sa

$$h_n = \mathbb{E}\left(\sum_{k=1}^{\infty} \xi_k^n\right). \tag{2.16}$$

Koristeći svojstvo (2.11) te nezavisnost od ξ i σ slijedi

$$\mathbb{E}\left(\sum_{k=1}^{\infty} \eta_k^n\right) = \mathbb{E}\left(\sigma^n \sum_{k=1}^{\infty} \xi_k^n\right) = \mathbb{E}(\sigma^n) h_n. \tag{2.17}$$

Budući da σ ima $\mathscr{G}(\lambda)$ razdiobu slijedi

$$\begin{aligned}
\mathbb{E}(\sigma^n) &= \int_0^{\infty} x^n \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x} dx \\
&= \int_0^{\infty} \frac{1}{\Gamma(\lambda)} x^{n+\lambda-1} e^{-x} dx \\
&= \frac{1}{\Gamma(\lambda)} \int_0^{\infty} x^{n+\lambda-1} e^{-x} dx \\
&= \frac{\Gamma(n+\lambda)}{\Gamma(\lambda)}.
\end{aligned}$$

Nadalje, prema Campbellovom teoremu slijedi

$$\mathbb{E}\left(\sum_{k=1}^{\infty} \eta_k^n\right) = \int_0^{\infty} z^n \lambda z^{-1} e^{-z} dz = \lambda \int_0^{\infty} z^{n-1} e^{-z} dz = \lambda \Gamma(n). \tag{2.18}$$

Uvrštavanjem odgovarajućih vrijednosti u jednakost (2.17) slijedi

$$h_n = \frac{\mathbb{E}\left(\sum_{k=1}^{\infty} \eta_k^n\right)}{\mathbb{E}(\sigma^n)} = \frac{\lambda\Gamma(n)}{\frac{\Gamma(n+\lambda)}{\Gamma(\lambda)}} = \frac{\lambda\Gamma(\lambda)\Gamma(n)}{\Gamma(n+\lambda)}. \tag{2.19}$$

Budući da vrijedi $\lambda\Gamma(\lambda) = \Gamma(\lambda+1)$, slijedi da je $(\lambda-1)\Gamma(\lambda-1) = \Gamma(\lambda)$. Uvrštavanjem u (2.19) slijedi

$$\begin{aligned} h_n &= \frac{\Gamma(\lambda+1)\Gamma(n)}{\Gamma(n+\lambda)} \\ &= \frac{\Gamma(\lambda+1)\Gamma(n)}{(n+\lambda-1)\Gamma(n+\lambda-1)} \\ &= \dots \\ &= \frac{\Gamma(\lambda+1)\Gamma(n)}{(\lambda+1)(\lambda+2)\dots(n+\lambda-1)\Gamma(\lambda+1)} \\ &= \frac{\Gamma(n)}{(1+\lambda)(2+\lambda)\dots(n+\lambda-1)}. \end{aligned}$$

Nadalje, kako je $n \in \mathbb{N}$ slijedi da je $\Gamma(n) = (n-1)!$. Stoga vrijedi

$$h_n = \frac{(n-1)!}{(\lambda+1)(\lambda+2)\dots(\lambda+n-1)}. \tag{2.20}$$

U prethodnom smo opisali specijalan slučaj rezultata poznatijeg kao Ewensova formula uzorkovanja (*eng. Ewens sampling formula*).

Ukoliko imamo nehomogen uzorak veličine n, tada sigurno postoji određen broj članova koji pripadaju jednoj vrsti i ne podudaraju se s ostalima. Uzmimo da ih ima a_1. Nadalje, uzmimo da imamo a_2 parova od kojih je svaki par istog tipa, ali različit od svih ostalih, a_3 trojki sa istim svojstvom i analogno dalje. Brojevi $a_1, \dots, a_n$ zadovoljavaju

$$a_1, \dots, a_n \geq 0, \qquad a_1 + 2a_2 + \dots + na_n = n, \tag{2.21}$$

i opisuju uzorak. Općenito, za svaki fiksni $n \geq 1$ definirajmo skup

$$\mathscr{A}_n = \left\{ (a_1, \dots, a_n) : a_k \geq 0, k = 1, \dots, n; \sum_{i=1}^{n} ia_i = n \right\}, \tag{2.22}$$

te uzmimo $\mathbf{A}_n = (a_1, \dots, a_n) \in \mathscr{A}_n$.

Ako $\mathbb{P}_n(\mathbf{A}_n = (a_1, \dots, a_n)) = \mathbb{P}_n(\mathbf{A}_n)$ označava vjerojatnost da je raspodjela vrsti u uzorku jednaka $(a_1, \dots, a_n)$, tada jednakost (2.20) predstavlja specijalan slučaj kada su svi pojedinci u uzorku iste vrste. To ćemo označiti sa $\mathbb{P}_n(n^1)$. Ewensova formula nam daje odgovarajuću vrijednost za općeniti slučaj. Kako bismo izračunali $\mathbb{P}_n(\mathbf{A}_n)$ primjetimo kako je vjerojatnost postizanja vrijednosti $(a_1, \ldots, a_n)$, uz dane populacijske frekvencije ξ_k (koje ćemo u nastavku označavati sa $\xi(k)$) jednaka

$$\begin{aligned}&\mathbb{P}_n\left(\mathbf{A}_n|\xi\right)\\ &= \frac{n!}{\prod (j!)^{a_j} a_j!} \sum \xi(k_{11})\xi(k_{12})\ldots\xi(k_{1a_1})\xi(k_{21})^2\ldots\xi(k_{2a_2})^2\ldots\xi(k_{n1})^n\ldots\xi(k_{na_n})^n,\end{aligned}$$

gdje sumaciju vršimo po različitim

$$k_{ij}, \qquad i = 1, \ldots, n, \qquad j = 1, \ldots, a_i.$$

Uvrštavanjem (2.11) u prethodnu jednakost slijedi

$$\begin{aligned}\mathbb{P}_n(\mathbf{A}_n|\xi) &= \frac{n!}{\prod (j!)^{a_j} a_j!} \frac{1}{\sigma^{1\cdot a_1 + 2\cdot a_2 + \cdots + n\cdot a_n}} \sum \eta(k_{11})\ldots\eta(k_{na_n})\\ &= \frac{n!}{\prod (j!)^{a_j} a_j!} \frac{1}{\sigma^n} \sum \eta(k_{11})\ldots\eta(k_{na_n}).\end{aligned}$$

Množenjem prethodne jednakosti sa σ^n dobivamo da je

$$\sigma^n \mathbb{P}_n(\mathbf{A}_n|\xi) = \frac{n!}{\prod (j!)^{a_j} a_j!} \sum \eta(k_{11})\ldots\eta(k^n_{na_n}). \tag{2.23}$$

Uzimanjem očekivanja u (2.23) slijedi

$$\mathbb{E}\left(\sigma^n \mathbb{P}_n(\mathbf{A}_n|\xi)\right) = \frac{n!}{\prod (j!)^{a_j} a_j!} \mathbb{E}\left(\sum \eta(k_{11})\ldots\eta(k^n_{na_n})\right).$$

Nezavisnost od σ i ξ povlači

$$\begin{aligned}\mathbb{E}(\sigma^n)\mathbb{E}\left(\mathbb{P}_n(\mathbf{A}_n|\xi)\right) &= \frac{n!}{\prod (j!)^{a_j} a_j!} \mathbb{E}\left(\sum \eta(k_{11})\ldots\eta(k^n_{na_n})\right),\\ \frac{\Gamma(n+\lambda)}{\Gamma(\lambda)}\mathbb{P}_n(a) &= \frac{n!}{\prod (j!)^{a_j} a_j!} \mathbb{E}\left(\sum \eta(k_{11})\ldots\eta(k^n_{na_n})\right).\end{aligned}$$

Potrebno je izračunati očekivanje na desnoj strani. Znamo da $\eta_i, i \in \mathbb{N}$ predstavljaju uređene visine skokova Moranovog procesa ϕ na segmentu $[0, \lambda]$ i da te visine formiraju Poissonov proces čija mjera očekivanja iznosi $\Lambda(dx) = \lambda z^{-1} e^{-z} dz$. Uzimanjem izmjerive funkcije $f : S \to \mathbb{R}, f_i(x) = x^i$ slijedi da očekivanje možemo zapisati kao

$$\mathbb{E}\left(\sum_{\text{različiti } Y_{ij} \in \mathscr{J}} f_1(Y_{11})\ldots f_1(Y_{1a_1}) f_2(Y_{2a_2})\ldots f_2(Y_{2a_2})\ldots f_n(Y_{na_n})\right), \tag{2.24}$$

gdje je $\mathscr{J}$ skup svih skokova procesa ϕ na intervalu $[0,\lambda]$. Koristeći Campbellov teorem, preciznije jednakost (1.23), slijedi

$$\mathbb{P}_n(\mathbf{A}_n) = \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)\prod(j!)^{a_j}a_j!}\mathbb{E}(f_1(Y_{11}))\dots\mathbb{E}(f_1(Y_{1a_1}))\mathbb{E}(f_2(Y_{2a_2}))\dots\mathbb{E}(f_2(Y_{2a_2}))\dots\mathbb{E}(f_n(Y_{na_n})).$$

Prema istom teoremu, preciznije jednakosti (1.26), slijedi da je

$$\begin{aligned}
\mathbb{P}_n(\mathbf{A}_n) &= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)\prod(j!)^{a_j}a_j!}\underbrace{\left(\int_S f_1(z)\Lambda(dz)\right)\dots\left(\int_S f_1(z)\Lambda(dz)\right)}_{a_1 \text{ puta}} \\
&\dots\underbrace{\left(\int_S f_n(z)\Lambda(dz)\right)\dots\left(\int_S f_n(z)\Lambda(dz)\right)}_{a_n \text{ puta}} \\
&= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)\prod(j!)^{a_j}a_j!}\left(\int_S f_1(z)\Lambda(dz)\right)^{a_1}\dots\left(\int_S f_n(z)\Lambda(dz)\right)^{a_n} \\
&= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)\prod(j!)^{a_j}a_j!}\prod_{j=1}^{n}\left(\lambda\int_S z^j z^{-1}e^{-z}dz\right)^{a_j} \\
&= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)}\prod_{j=1}^{n}\frac{1}{j!^{a_j}a_j!}\left(\lambda\int_S z^{j-1}e^{-z}dz\right)^{a_j} \\
&= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)}\prod_{j=1}^{n}\frac{1}{j!^{a_j}a_j!}(\lambda\Gamma(j))^{a_j} \\
&= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)}\prod_{j=1}^{n}\frac{\lambda^{a_j}}{a_j!}\left(\frac{(j-1)!}{j!}\right)^{a_j} \\
&= \frac{\Gamma(\lambda)n!}{\Gamma(n+\lambda)}\prod_{j=1}^{n}\frac{1}{a_j!}\left(\frac{\lambda}{j}\right)^{a_j}.
\end{aligned}$$

Bibliografija

[1] Shui Feng, *The Poisson-Dirichlet distribution and related topics: models and asymptotic behaviors*, Springer, 2010.

[2] John Frank Charles Kingman, *Poisson Processes*, Clarendon Press Oxford, 2002.

Sažetak

Ovaj diplomski rad bavi se Poisson - Dirichletovom razdiobom te njenim glavnim teorijskim svojstvima. Pojednostavljeno, možemo reći da se Poisson - Dirichlet razdioba dobije kao granična razdioba Dirichletove razdiobe, odnosno, grubo govoreći, uređajne statistike slučajnih vektora sa Dirichletovom razdiobom $\mathscr{D}(\alpha_1, \ldots, \alpha_n)$ teže prema Poisson - Dirichletovoj razdiobi kada $n \to \infty$ i $\sum_{k=1}^{n} \alpha_n \to \lambda$. Nadalje, u radu su prezentirane i formule za granične razdiobe: formule za distribuciju i zajedničku gustoću prvih r komponenti slučajnog vektora $(\xi_1, \xi_2, \ldots) \sim \mathscr{P}\mathscr{D}(\lambda)$. Rad smo zaključili sa jednim poznatim rezultatom, uključujući i izvod. Riječ je o *Ewensovoj formuli uzorkovanja* koja se koristi za opis ravnotežne distribucije određenih difuzijskih modela u populacijskoj genetici (Za detalje v. [1]). U radu su također prezentirani osnovni pojmovi potrebni za razumijevanje središnjeg djela rada. Uveli smo definiciju Poissonovog procesa te iskaze i dokaze određenih teorema relevantnih za naša razmatranja. Također smo uveli pojam subordinatora koji je definiran kao monoton slučajni proces sa nezavisnim i stacionarnim prirastima, i naveli smo jedan poseban primjer subordinatora poznatiji pod imenom Moranov subordinator, koji ima važnu ulogu u konstrukciji Poisson -Dirichletovog procesa.

Životopis

Ja, Vanja Jović, autor ovog rada, sam rođen 31.01.1990.g. u Doboju, BiH. Svoje prve godine života sam proveo u Njemačkoj gdje sam započeo osnovnu školu i dovršio prva dva razreda. Osnovnu školu sam završio u Zagrebu, gdje sam se doselio sa osam godina. U ljeto 2004.g. upisujem XIII. Gimnaziju u Zarebu, smjer prirodoslovno - matematički koju sam završio 2008.g. Nakon mature upisujem studij preddiplomske matematike na Prirodoslovno - Matematičkom Fakultetu Matematički odsjek, a u ljeto 2012.g. upisujem diplomski studij Financijske i poslovne matematike. Od izvannastavnih aktivnosti vrijedi istaknuti dvije ljetne prakse koje sam odradio u Njemačkoj te školu stranih jezika koju sam pohađao paralelno uz studij. Tečno govorim engleski i njemački jezik.