Bayesovska statistika i Dirichletov proces

Ćuk, Damir

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Damir Ćuk

BAYESOVSKA STATISTIKA I DIRICHLETOV PROCES

Diplomski rad

Voditelj rada:
prof. dr. sc. Bojan Basrak

Zagreb, rujan, 2022.

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. ______________________________

Profesorici Danici Lukač, za iznimno vrijedne matematičke temelje.
Roditeljima, za podršku tijekom obrazovanja.

Sadržaj

Uvod

U ovom radu uvodimo pojam *Dirichletovog procesa* u kontekstu *bayesovske statistike*, po uzoru na [1] i [2]. Dirichletov proces, kojeg je prvi definirao Thomas S. Ferguson (1929. - danas, [1], 1973.), usko je vezan uz *Dirichletovu distribuciju*, koja je pak svoj naziv dobila po njemačkom matematičaru Peteru Gustavu Lejeuneu Dirichletu (1805. - 1859.). Usprkos nekim nedostacima, Dirichletov se proces pokazao korisnim u raznim problemima strojnoga učenja.

Dirichletov proces pripada klasi slučajnih procesa koje nazivamo slučajnim mjerama, preciznije slučajnim vjerojatnosnim mjerama. Slično kao što je to učinio Ferguson, Dirichletov proces definirat ćemo pomoću njegovih (marginalnih) konačno-dimenzionalnih distribucija, a opravdanost definicije slijedit će iz Kolmogorovljevih uvjeta suglasnosti.

U poglavlju 1 prisjećamo se osnovnih pojmova teorije vjerojatnosti koji će nam biti potrebni u nastavku rada. Budući se veliki dio bayesovske statistike svodi na računanje s funkcijama gustoća slučajnih varijabli i vektora, precizno smo definirali te pojmove, čak i za jedan donekle nestandardan slučaj. Nakon toga uvodimo pojam slučajnoga procesa te opisujemo način na koji se on može zadati pomoću suglasne familije konačno-dimenzionalnih distribucija. Definicije su potkrijepljene primjerima, a sve su tvrdnje ili dokazane ili je čitatelj upućen na literaturu u kojoj se dokaz može pronaći. Međutim, pretpostavlja se da čitatelju ovo ipak nije prvi susret s pojmovima koji su uvedeni u poglavlju 1.

U poglavlju 2 upoznajemo se s osnovnom idejom bayesovskoga zaključivanja, tzv. *parametarskom bayesovskom statistikom*. Precizno je definiran pojam uvjetne funkcije gustoće, pokazana je Bayesova formula te su objašnjeni pojmovi *apriorne* i *aposteriorne* gustoće. Teorija je ukupno potkrijepljena sa tri primjera, od kojih je jedan izložen kao motivacijski.

U poglavlju 3 proširujemo ideju iz poglavlja 2 u tzv. *neparametarsku bayesovsku statistiku*. Uvjetnu funkciju gustoće ovdje zamjenjuje slučajna vjerojatnost, odnosno slučajna vjerojatnosna mjera pa je definiran i taj pojam. Uvedena je Dirichletova distribucija, a Dirichletov proces definiran je kao ona slučajna vjerojatnost čije su konačno-dimenzionalne distribucije upravo Dirichletove. Definicija je opravdana provjerom Kolmogorovljevih uvjeta suglasnosti. Konačno, primjerima je pokazano kako Dirichletov proces koristimo u neparametarskoj bayesovskoj statistici te je napravljena mala simulacijska studija.

Poglavlje 1

Osnovni pojmovi teorije vjerojatnosti

U ovom poglavlju uvodimo osnovne pojmove i rezultate teorije vjerojatnosti koji će nam biti potrebni u nastavku rada. To činimo po uzoru na [5]. Iskazane tvrdnje uglavnom neće biti dokazivane, već se dokazi mogu pronaći u [5]. Osnovne pojmove i rezultate teorije mjere i integrala nećemo formalno uvoditi, ali će se oni koristiti onako kako su uvedeni u [4].

1.1 Slučajne varijable

Sa $\mathcal{B}$ ćemo označavati Borelovu σ-algebru na $\mathbb{R}$. Podsjetimo se da funkciju $X : \Omega \to \mathbb{R}$ zovemo **slučajnom varijablom** na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ ako je ona $(\mathcal{F}, \mathcal{B})$-izmjeriva. Slično, za $n \in \mathbb{N}$, sa $\mathcal{B}_n$ označavamo Borelovu σ-algebru na $\mathbb{R}^n$ i kažemo da je funkcija $\mathbb{X} : \Omega \to \mathbb{R}^n$ **slučajni vektor** na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ ako je ona $(\mathcal{F}, \mathcal{B}_n)$-izmjeriva.
Pokazuje se da je $\mathbb{X} : \Omega \to \mathbb{R}^n$, $\mathbb{X} = (X_1, ..., X_n)$ slučajni vektor na $(\Omega, \mathcal{F}, \mathbb{P})$ ako i samo ako je $X_k : \Omega \to \mathbb{R}$ slučajna varijabla na $(\Omega, \mathcal{F}, \mathbb{P})$ za svako $k = 1, ..., n$. Dakle, slučajni vektori su uređene n-torke slučajnih varijabli. Slučajnu varijablu možemo smatrati specijalnim oblikom slučajnoga vektora za $n = 1$.

Definicija 1.1.1. *Za slučajnu varijablu X na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ kažemo da je **diskretna slučajna varijabla** ako postoji konačan ili prebrojiv skup $D \subset \mathbb{R}$ takav da je $\mathbb{P}(X \in D) = 1$. U tom slučaju definiramo **diskretnu funkciju gustoće** od X kao funkciju $f_X : \mathbb{R} \to \mathbb{R}$*

$$f_X(x) = \mathbb{P}(X = x),\ x \in \mathbb{R}.$$

Nije teško vidjeti da se za zadani konačan ili prebrojiv skup $D = \{a_1, a_2, ...\} \subset \mathbb{R}$ i niz $(p_n)_{n\in\mathbb{N}} \subset [0, 1]$ takav da je $\sum_{n=1}^{\infty} p_n = 1$, može konstruirati (neki) vjerojatnosni prostor i diskretna slučajna varijabla X na tom vjerojatnosnom prostoru sa diskretnom funkcijom gustoće $f_X : \mathbb{R} \to \mathbb{R}$, za koju vrijedi

$$f_X(x) = \mathbb{P}(X = x) = \begin{cases} p_k, & \text{ako je } x = a_k \text{ za neko } k \in \mathbb{N} \\ 0, & \text{inače.} \end{cases}$$

(Ukoliko je $D = \{a_1, ..., a_n\}$ pretpostavljamo da je $p_k = 0,\ k > n$.) U tom ćemo slučaju reći da je distribucija diskretne slučajne varijable X zadana diskretnom funkcijom gustoće f_X, odnosno da je dana tablicom

$$X \sim \left(\begin{array}{cccc} a_1 & a_2 & a_3 & ... \\ p_1 & p_2 & p_3 & ... \end{array} \right),$$

pri čemu nije važno nad kojim je točno vjerojatnosnim prostorom ta slučajna varijabla zadana.

Definicija 1.1.2. *Za slučajnu varijablu X na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$ *definiramo njezinu* ***funkciju distribucije*** $F_X : \mathbb{R} \to \mathbb{R}$ *sa*

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in \langle -\infty, x]),\ x \in \mathbb{R}.$$

Ukoliko je X diskretna slučajna varijabla, uz oznake kao ranije, lagano se vidi da vrijedi

$$F_X(x) = \sum_{y \leq x} f_X(y)\,,\ x \in \mathbb{R},$$

gdje je gornja suma najviše prebrojiva, budući je $f_X(y) = 0$ za $y \notin D$. Zapravo sumiramo po svim $y \in D \cap \langle -\infty, x]$. Štoviše, čim je zadano f_X, za proizvoljan $B \in \mathcal{B}$ možemo računati vjerojatnosti oblika

$$\mathbb{P}(X \in B) = \sum_{x \in B} f_X(x). \tag{1.1}$$

Definicija 1.1.3. *Za slučajnu varijablu X na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ kažemo da je **neprekidna slučajna varijabla** ako postoji nenegativna Borelova funkcija $f_X : \mathbb{R} \to \mathbb{R}$ takva da je*

$$F_X(x) = \int_{\langle -\infty, x]} f_X \, d\lambda, \ x \in \mathbb{R},$$

*gdje su F_X funkcija distribucije slučajne vraijable X, a λ Lebesgueova mjera na $(\mathbb{R}, \mathcal{B})$. Takvu funkciju f_X tada nazivamo **funkcijom gustoće neprekidne slučajne varijable** X.*

Integral u prethodnoj definiciji, kao i u idućem teoremu, je Lebesgueov integral po Lebesgueovoj mjeri λ (za detalje v. [4]).

Teorem 1.1.4. *([5], Propozicija 9.5.) Neka je $f : \mathbb{R} \to \mathbb{R}$ Borelova funkcija. Da bi f bila funkcija gustoće (neke) neprekidne slučajne varijable (na nekom vjerojatnosnom prostoru) nužno je i dovoljno da vrijedi*

$$(1) \ \ f(x) \geq 0, \ x \in \mathbb{R},$$

$$(2) \ \ \int_{\mathbb{R}} f \, d\lambda = 1.$$

Dakle, ako je dana Borelova funkcija $f : \mathbb{R} \to \mathbb{R}$ koja zadovoljava svojstva *(1)* i *(2)* iz prethodnog teorema, kažemo da ona po distribuciji zadaje neku neprekidnu slučajnu varijablu X čija je f funkcija gustoće. Pritom ponovno nije važno nad kojim je točno vjerojatnosnim prostorom ta slučajna varijabla zadana. Slično kao u (1.1), za neprekidnu slučajnu varijablu X s funkcijom gustoće f_X i $B \in \mathcal{B}$, možemo računati vjerojatnosti oblika

$$\mathbb{P}(X \in B) = \int_B f_X \, d\lambda. \tag{1.2}$$

U nastavku ovoga rada sve će slučajne varijable biti isključivo diskretne ili neprekidne. Općenito ćemo za slučajnu varijablu X reći da ima gustoću, odnosno funkciju gustoće, pri čemu će se mislti na diskretnu funkciju gustoće, ukoliko je X diskretna, odnosno na funkciju gustoće iz definicije 1.1.3, ukoliko je X neprekidna slučajna varijabla. Uglavnom nećemo precizirati na kojem je vjerojatnosnom prostoru X zadana, budući nam je poznavanje njezine funkcije gustoće dovoljno da bismo računali vjerojatnosti oblika (1.1), odnosno

(1.2). Umjesto izraza *distribucija*, koristi se i izraz *razdioba*. Ukoliko istovremeno uvodimo više različitih slučajnih varijabli i vektora, podrazumijevamo da su ti objekti zadani na istom vjerojatnosnom prostoru.

Podsjetimo se i pojma **(matematičkog) očekivanja** slučajne varijable X. Ono se definira kao Lebesgueov integral

$$\mathbb{E}X = \int_\Omega X \, \mathrm{d}\mathbb{P},$$

čim takav postoji (može biti i $\mathbb{E}X = \pm\infty$). Kao i ranije, pokazuje se da je za računanje očekivanja diskretne ili neprekidne slučajne varijable X dovoljno znati njezinu gustoću f_X. Ukoliko se radi o diskretnoj slučajnoj varijabli, imamo

$$\mathbb{E}X = \sum_{x\in\mathbb{R}} x \cdot f_X(x), \tag{1.3}$$

čim gornja suma konvergira ili divergira u $\pm\infty$. Ukoliko je X neprekidna slučajna varijabla, imamo

$$\mathbb{E}X = \int_\mathbb{R} x \cdot f_X(x) \ \mathrm{d}\lambda(x), \tag{1.4}$$

čim takav integral postoji.

Napomena 1.1.5. ([4], Dodatak B) Neka je $B \in \mathcal{B}$ Borelov skup u $\mathbb{R}$ po kojem ima smisla računati (nepravi) Riemannov integral (npr. B je konačna ili prebrojiva unija intervala) te neka je $f : \mathbb{R} \to \mathbb{R}$ Riemann-integrabilna Borelova funkcija. Pokazuje se da je f tada i Lebesgue integrabilna te da vrijedi

$$\int_B f \, \mathrm{d}\lambda = \int_B f(x) \, \mathrm{d}x,$$

odnosno da se Lebesgueov i Riemannov integral podudaraju.

U našem će se radu na sve probleme integriranja moći primijeniti tvrdnja prethodne napomene, tako da ćemo umjesto Lebesgueovih integrala zapravo računati Riemannove. Precizniji iskaz gornje tvrdnje, zajedno s dokazom, može se pronaći u [4], dodatak B.

Za kraj navedimo primjer dvije slučajne varijable koje ćemo koristiti u nastavku.

Primjer 1.1.6. Za diskretnu slučajnu varijablu X kažemo da je **Bernoullijeva s parametrom** p, $p \in [0, 1]$, ukoliko je

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}.$$

Pišemo $X \sim B(p)$. Koristeći (1.3) dobivamo

$$\mathbb{E}(X) = 0 \cdot (1-p) + 1 \cdot p = p.$$

Prisjetimo se funkcije $B : \langle 0, \infty\rangle \times \langle 0, \infty\rangle \to \langle 0, \infty\rangle$ dane sa

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt\,, \quad a, b > 0, \tag{1.5}$$

koju zovemo **beta funkcijom**. Također, funkciju $\Gamma : \langle 0, \infty\rangle \to \langle 0, \infty\rangle$ danu sa

$$\Gamma(a) = \int_0^\infty e^{-t} t^{a-1}\, dt\,, \quad a > 0, \tag{1.6}$$

zovemo **gama funkcijom**.

Lema 1.1.7. *Za beta i gama funkcije, dane sa* (1.5)*, odnosno* (1.6)*, vrijedi*

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{1.7}$$

za sve $a, b > 0$.

Dokaz. ([5], Primjer 9.12.) Za $a, b > 0$ imamo

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt = \left\{ t = \sin^2 \phi \right\} =$$

$$= 2 \int_0^{\pi/2} (\sin \phi)^{2a-1} (\cos \phi)^{2b-1}\, d\phi.$$

S druge strane je

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\, dt = \left\{ t = s^2 \right\} = 2 \int_0^\infty s^{2a-1} e^{-s^2}\, ds$$

pa imamo

$$\begin{aligned}
\Gamma(a)\Gamma(b) &= 4\left(\int_0^\infty s^{2a-1}e^{-s^2}\,ds\right)\left(\int_0^\infty t^{2b-1}e^{-t^2}\,dt\right) = \\
&= 4\int_0^\infty \int_0^\infty s^{2a-1}t^{2b-1}e^{-(s^2+t^2)}\,ds\,dt = \\
&= \left\{t = r\cos\phi,\ s = r\sin\phi\right\} = \\
&= 4\left(\int_0^\infty r^{2(a+b)-1}e^{-r^2}\,dr\right)\left(\int_0^\infty (\sin\phi)^{2a-1}(\cos\phi)^{2b-1}\,d\phi\right) = \\
&= \Gamma(a+b)\cdot B(a,b).
\end{aligned}$$

Odavde, zbog $\Gamma(a+b) > 0$ slijedi (1.7).

□

Lema 1.1.8. *Za Gamma funkciju danu sa* (1.6) *vrijedi*

$$\Gamma(a+1) = a\cdot\Gamma(a) \tag{1.8}$$

za svako $a > 0$.

Dokaz. Koristeći relaciju (1.7) dobivamo

$$\Gamma(a+1) = \frac{\Gamma(a)\Gamma(1)}{B(a,1)}.$$

Pri tome su

$$\Gamma(1) = \int_0^\infty e^{-t}\cancel{t^{1-1}}\,dt = [-e^{-t}]_{t=0}^\infty = 1$$

te

$$B(a,1) = \int_0^1 t^{a-1}\cancel{(1-t)^{1-1}}\,dt = \left[\frac{t^a}{a}\right]_{t=0}^1 = \frac{1}{a}.$$

Stoga vrijedi (1.8).

□

Primjer 1.1.9. Za neprekidnu slučajnu varijablu X kažemo da ima **beta razdiobu s parametrima** a i b, $a, b > 0$, ukoliko je

$$f_X(x) = \begin{cases} \frac{1}{B(a,b)} \cdot x^{a-1}(1-x)^{b-1}, & 0 \leq x \leq 1 \\ 0, & \text{inače,} \end{cases}$$

gdje je $B(a, b)$, dano sa (1.5), normalizirajuća konstanta. Funkcija f_X očito zadovoljava svojstva (1) i (2) iz teorema 1.1.4. Pišemo $X \sim Beta(a, b)$. Osnovnu primjenu ovakve distribucije u bayesovskoj statistici prikazat ćemo u poglavlju 2, dok će u poglavlju 3 igrati važnu ulogu prilikom definicije Dirichletovog procesa. Koristeći (1.4) i napomenu 1.1.5 računamo

$$\mathbb{E}(X) = \int_0^1 x \cdot \frac{1}{B(a,b)} \cdot x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$$

$$= \frac{1}{B(a,b)} \cdot \int_0^1 x^a(1-x)^{b-1}\,\mathrm{d}x = \frac{B(a+1,b)}{B(a,b)}.$$

Koristeći (1.7) i (1.8) dobivamo

$$\mathbb{E}(X) = \frac{\Gamma(a+1)\cancel{\Gamma(b)} \cdot \Gamma(a+b)}{\Gamma(a+1+b) \cdot \Gamma(a)\cancel{\Gamma(b)}} = \frac{a \cdot \cancel{\Gamma(a)}\cancel{\Gamma(a+b)}}{(a+b) \cdot \cancel{\Gamma(a+b)}\cancel{\Gamma(a)}} = \frac{a}{a+b}.$$

1.2 Slučajni vektori. Nezavisnost

Analogno pojmovima diskretne i neprekidne slučajne varijable, uvodimo pojmove diskretnog i neprekidnog slučajnog vektora. U čitavom odjeljku uzimamo da je $n \in \mathbb{N}$.

Definicija 1.2.1. *Za slučajni vektor $\mathbb{X} : \Omega \to \mathbb{R}^n$ na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ kažemo da je **diskretan slučajni vektor** ako postoji konačan ili prebrojiv $D \subset \mathbb{R}^n$ takav da je $\mathbb{P}(\mathbb{X} \in D) = 1$. Funkciju $f_{\mathbb{X}} : \mathbb{R}^n \to \mathbb{R}$ za koju vrijedi*

$$f_{\mathbb{X}}(\boldsymbol{x}) = \mathbb{P}(\mathbb{X} = \boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^n$$

*tada nazivamo **diskretnom funkcijom gustoće** slučajnog vektora $\mathbb{X}$.*

Slično kao u (1.1), za diskretan slučajni vektor $\mathbb{X}$ i $B \in \mathcal{B}_n$, računat ćemo

$$\mathbb{P}(\mathbb{X} \in B) = \sum_{\mathbf{x} \in B} f_{\mathbb{X}}(\mathbf{x}), \tag{1.9}$$

gdje je gornja suma ponovno najviše prebrojiva, budući je $f_{\mathbb{X}}(\mathbf{x}) = 0$ za $\mathbf{x} \notin D$.

Definicija 1.2.2. *Neka je* $\mathbb{X} = (X_1, ..., X_n)$ *slučajni vektor na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$. *Funkciju* $F_{\mathbb{X}} : \mathbb{R}^n \to \mathbb{R}$, *danu sa*

$$F_{\mathbb{X}}(\boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, ..., X_n \leq x_n), \quad \boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$$

nazivamo ***funkcijom distribucije slučajnog vektora*** $\mathbb{X}$.

Definicija 1.2.3. *Za slučajni vektor* $\mathbb{X} = (X_1, ..., X_n)$ *na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$ *kažemo da je* ***neprekidan slučajni vektor*** *ako postoji nenegativna Borelova funkcija* $f_{\mathbb{X}} : \mathbb{R}^n \to \mathbb{R}$ *takva da je*

$$F_{\mathbb{X}}(\boldsymbol{x}) = \int_{\langle -\infty, \boldsymbol{x}]} f_{\mathbb{X}} \, d\lambda^n, \; \boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n,$$

gdje je $\langle -\infty, \boldsymbol{x}] = \langle -\infty, x_1] \times ... \times \langle -\infty, x_n]$. *Takvu funkciju* $f_{\mathbb{X}}$ *tada nazivamo* ***funkcijom gustoće neprekidnog slučajnog vektora*** $\mathbb{X}$.

Slično kao u (1.2), za neprekidan slučajni vektor $\mathbb{X}$ i $B \in \mathcal{B}_n$, računat ćemo

$$\mathbb{P}(\mathbb{X} \in B) = \int_B f_{\mathbb{X}} \, d\lambda^n. \tag{1.10}$$

Nije teško pokazati da je slučajni vektor $\mathbb{X} = (X_1, ..., X_n)$ diskretan ako i samo ako su pripadne slučajne varijable $X_1, ..., X_n$ diskretne. Analogna tvrdnja za neprekidne slučajne varijable općenito ne vrijedi. Ukoliko su $X_1, ..., X_n$ neprekidne slučajne varijable, slučajni vektor $\mathbb{X} = (X_1, ..., X_n)$ nije nužno neprekidan.

Primjer 1.2.4. Neka je X neprekidna slučajna varijabla. Kada bi slučajni vektor $\mathbb{X} = (X, X)$ bio neprekidan s funkcijom gustoće $f_{\mathbb{X}}$, tada bi za Borelov skup $B = \{(x, x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$ vrijedilo

$$1 = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}((X, X) \in B) = \int_B f_{\mathbb{X}} \, d\lambda^2 = 0,$$

jer je $\lambda^2(B) = 0$.

Osim diskretnih i neprekidnih, koristit ćemo još jednu specijalnu vrstu slučajnih vektora. Uzimamo $n, m \in \mathbb{N}$.

Definicija 1.2.5. *Neka je $\mathbb{X} = (X_1, ..., X_n)$ diskretan, a $\mathbb{Y} = (Y_1, ..., Y_m)$ neprekidan slučajni vektor na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Za nenegativnu Borelovu funkciju $f_{\mathbb{X},\mathbb{Y}} : \mathbb{R}^{n+m} \to \mathbb{R}$ kažemo da je gustoća slučajnog vektora $(\mathbb{X}, \mathbb{Y}) = (X_1, ..., X_n, Y_1, ..., Y_m)$, ako za sve $B_n \in \mathcal{B}_n$ i $B_m \in \mathcal{B}_m$ vrijedi*

$$\mathbb{P}(\mathbb{X} \in B_n, \mathbb{Y} \in B_m) = \sum_{\boldsymbol{x} \in B_n} \int_{B_m} f_{\mathbb{X},\mathbb{Y}}(\boldsymbol{x}, \boldsymbol{y})\, d\lambda^m(\boldsymbol{y}). \tag{1.11}$$

Jasno je da analognu gustoću možemo definirati i za slučajni vektor $(\mathbb{X}, \mathbb{Y})$, gdje je $\mathbb{X}$ neprekidan, a $\mathbb{Y}$ diskretan slučajni vektor. Tu smo gustoću definirali motivirani upravo jednakostima (1.9) i (1.10). Uočimo da za diskretan slučajni vektor $\mathbb{X}$ i neprekidan slučajni vektor $\mathbb{Y}$ vrijedi

$$f_{\mathbb{X}}(\mathbf{x}) = \mathbb{P}(\mathbb{X} = \mathbf{x}) = \mathbb{P}(\mathbb{X} = \mathbf{x}, \mathbb{Y} \in \mathbb{R}) \geq \mathbb{P}(\mathbb{X} = \mathbf{x}, \mathbb{Y} \in B_m) = \int_{B_m} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})\, d\lambda^m(\mathbf{y}), \tag{1.12}$$

za svako $\mathbf{x} \in \mathbb{R}^n$ i $B_m \in \mathcal{B}_m$. Kako je već $f_{\mathbb{X}}(\mathbf{x})$ različito od nula za najviše prebrojivo mnogo $\mathbf{x} \in \mathbb{R}^n$, to je i suma s desne strane u (1.11) najviše prebrojiva. Također, budući je $f_{\mathbb{X},\mathbb{Y}} \geq 0$, iz Lebesgueovog teorema o monotonoj konvergenciji (za detalje v. [4]) imamo

$$\sum_{\mathbf{x} \in B_n} \int_{B_m} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})\, d\lambda^m(\mathbf{y}) = \int_{B_m} \sum_{\mathbf{x} \in B_n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})\, d\lambda^m(\mathbf{y}).$$

U nastavku rada ćemo promatrati slučajne vektore $\mathbb{X} = (X_1, ..., X_n)$ i $\mathbb{Y} = (Y_1, ..., Y_m)$ koji su diskretni ili neprekidni, a $f_{\mathbb{X}}$ i $f_{\mathbb{Y}}$ njima pripadne funkcije gustoća (iz definicije 1.2.1 ili definicije 1.2.3). Nadalje, za slučajni vektor $(\mathbb{X}, \mathbb{Y}) = (X_1, ..., X_n, Y_1, ..., Y_m)$ ćemo pretpostavljati da ima funkciju gustoće $f_{\mathbb{X},\mathbb{Y}}$, iz definicije 1.2.1, definicije 1.2.3 ili definicije 1.2.5). Specijalno, ukoliko su $\mathbb{X}$ i $\mathbb{Y}$ neprekidni slučajni vektori, dodatno pretpostavljamo da je to i slučajni vektor $(\mathbb{X}, \mathbb{Y})$.

Napomena 1.2.6. Općenito govoreći, *zadati distribuciju slučajnoga vektora* zapravo znači zadati njegovu funkciju distribucije iz definicije 1.2.2. Nije teško pokazati (v. [5], 9.3.) da za takvu funkciju $F_{\mathbb{X}}$ vrijede sljedeća četiri svojstva:

(1) $F_{\mathbb{X}}$ je (po komponentama) neopadajuća, tj. $(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n)\ \mathbf{x} \leq \mathbf{y} \implies F_{\mathbb{X}}(\mathbf{x}) \leq F_{\mathbb{X}}(\mathbf{y})$,

(2) $F_{\mathbb{X}}$ je (po komponentama) zdesna neprekidna, tj. $(\forall \mathbf{x}_0 \in \mathbb{R}^n)\ \lim\limits_{\mathbf{x} \searrow \mathbf{x}_0} F_{\mathbb{X}}(\mathbf{x}) = F_{\mathbb{X}}(\mathbf{x}_0)$,

(3) $F_{\mathbb{X}}(x_1, ..., x_n) \to 0$, ako $x_k \to -\infty$ za barem jedno $k = 1, ..., n$ i

(4) $F_{\mathbb{X}}(x_1, ..., x_n) \to 1$, ako $x_k \to \infty$ za svako $k = 1, ..., n$.

Funkcije $F : \mathbb{R}^n \to \mathbb{R}$ koje zadovoljavaju gornja svojstva (1)-(4) nazivamo **vjerojatnosnim (konačno-dimenzionalnim) funkcijama distribucija**. Pokazuje se ([5], 9.3.) da za svaku takvu funkciju $F : \mathbb{R}^n \to \mathbb{R}$ postoje vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbb{P})$ i slučajni vektor $\mathbb{X} : \Omega \to \mathbb{R}^n$ na njemu čija je F funkcija distribucije. Jasno je i da smo zadavanjem funkcije gustoće slučajnog vektora (na jedan od tri prethodno opisana načina; v. definicije 1.2.1, 1.2.3, 1.2.5) ujedno zadali i njegovu funkciju distribucije.

Podsjetimo se sada još jednoga važnog pojma, a to je nezavisnost. Za slučajne varijable $X_1, ..., X_n$ na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ kažemo da su **nezavisne**, ako za proizvoljne $B_1, ..., B_n \in \mathcal{B}$ vrijedi

$$\mathbb{P}(X_1 \in B_1, ..., X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

Lako se vidi da su diskretne slučajne varijable $X_1, ..., X_n$ nezavisne ako i samo ako za svako $(x_1, ..., x_n) \in \mathbb{R}^n$ vrijedi

$$f_{\mathbb{X}}(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i). \tag{1.13}$$

Za neprekidne slučajne varijable imamo sličnu tvrdnju.

Teorem 1.2.7. *([5], Teorem 11.2.) Neka je dan vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbb{P})$. Ako je slučajni vektor $\mathbb{X} = (X_1, ..., X_n)$ neprekidan, odnosno ima funkciju gustoće $f_{\mathbb{X}}$ iz definicije 1.2.3, tada su sve slučajne varijable $X_1, ..., X_n$ nužno neprekidne, odnosno imaju funkcije gustoće $f_{X_1}, ..., f_{X_n}$ respektivno, iz definicije 1.1.3. (Obrat ove tvrdnje općenito ne vrijedi, v. primjer 1.2.4). Osim toga, u tom su slučaju $X_1, ..., X_n$ nezavisne ako i samo ako vrijedi* (1.13)*, za svaki $(x_1, ..., x_n) \in \mathbb{R}^n$, osim eventualno na Borelovom podskupu od $\mathbb{R}^n$ Lebesgueove mjere nula.*

1.3 Slučajni procesi

U ovom odjeljku uvodimo posljednji veliki pojam teorije vjerojatnosti koji ćemo koristiti u ovome radu, a to je slučajni proces. Taj će nam pojam biti potreban u poglavlju 3.

U čitavom odjeljku uzimamo da je $n \in \mathbb{N}$ te da je T proizvoljan skup indeksa. Sa $\mathbb{R}^T$ označavamo skup svih funkcija $x : T \to \mathbb{R}$. Za vrijednost koju funkcija $x \in \mathbb{R}^T$ poprima u točki $t \in T$ ćemo koristiti uobičajeniju oznaku x_t, umjesto $x(t)$. Iduće što želimo je definirati neku σ-algebru na $\mathbb{R}^T$. To činimo po uzoru na [5], 9.4.

Neka je $\{t_1, ..., t_n\}$ proizvoljan konačni podskup od T. Definiramo **projekciju** $\pi_{t_1,...,t_n} : \mathbb{R}^T \to \mathbb{R}^n$ **nad koordinatama** $t_1, ..., t_n$ sa

$$\pi_{t_1,...,t_n}(x) = (x_{t_1}, ..., x_{t_n}), \quad x \in \mathbb{R}^T.$$

Definicija 1.3.1. *Za skup $A \subseteq \mathbb{R}^T$ kažemo da je **cilindar s bazom** M **nad koordinatama** $t_1, ..., t_n$ ako postoje neprazan konačni podskup $\{t_1, ..., t_n\}$ od T i skup $M \subseteq \mathbb{R}^n$ takvi da je*

$$A = \pi^{-1}_{t_1,...,t_n}(M) = \left\{ x \in \mathbb{R}^T : (x_{t_1}, ..., x_{t_n}) \in M \right\}.$$

*Ako je $M \in \mathcal{B}_n$, tj. Borelov skup u $\mathbb{R}^n$, tada za A kažemo da je **Borelov cilindar**. Skup svih Borelovih cilindara u $\mathbb{R}^T$ označavamo sa $\mathcal{F}_T$. **Borelovom σ-algebrom na** $\mathbb{R}^T$ zovemo σ-algebru generiranu sa $\mathcal{F}_T$ i označavamo sa $\mathcal{B}_T = \sigma(\mathcal{F}_T)$. Elemente od $\mathcal{B}_T$ nazivamo **Borelovim skupovima** u $\mathbb{R}^T$.*

Definicija 1.3.2. *Neka je dan vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbb{P})$ i T neki skup indeksa. Reći ćemo da je funkcija $\mathbb{X} : \Omega \to \mathbb{R}^T$ **slučajni proces**, ako je ona $(\mathcal{F}, \mathcal{B}_T)$-izmjeriva.*

Primjer 1.3.3. Neka je $\Omega = \{-1, 1\}$ te neka su $\mathcal{F}$ partitivni skup od Ω (skup svih podskupova od Ω) i $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ vjerojatnost na $(\Omega, \mathcal{F})$ za koju vrijedi

$$\mathbb{P}(\{-1\}) = \mathbb{P}(\{1\}) = \frac{1}{2}.$$

Neka je $T = \mathbb{N}$ te $\mathbb{X} : \Omega \to \mathbb{R}^{\mathbb{N}}$ zadana sa

$$\mathbb{X}(\omega)_n = \mathbb{X}(\omega)(n) = \omega \cdot n, \quad \omega \in \Omega, \ n \in \mathbb{N}.$$

Ta je funkcija $(\mathcal{F}, \mathcal{B}_T)$-izmjeriva, budući $\mathcal{F}$ sadrži sve moguće podskupove od Ω. Štoviše, nije teško vidjeti da za proizvoljno $A \in \mathcal{B}_T$ vrijedi

$$\mathbb{X}^{-1}(A) = \begin{cases} \emptyset, & -x, x \notin A \\ \{-1\}, & -x \in A, x \notin A \\ \{1\}, & -x \notin A, x \in A \\ \Omega, & -x, x \in A \end{cases},$$

gdje je $x : \mathbb{N} \to \mathbb{R}$, dana sa

$$x_n = x(n) = n, \quad n \in \mathbb{N}.$$

Sjetimo se da su elementi od $\mathcal{B}_T$ podskupovi od $\mathbb{R}^T = \mathbb{R}^{\mathbb{N}}$ pa se za $A \in \mathcal{B}_T$ ima smisla pitati jesu li $-x, x \in A$. Dakle, $\mathbb{X}$ je slučajni proces. Općenito, slučajne procese za koje je $T = \mathbb{N}$ nazivamo i **slučajnim nizovima**. Možemo reći da $\mathbb{X}$ poprima oblike $x = (1, 2, 3, ...)$ i $-x = (-1, -2, -3, ...)$ sa jednakim vjerojatnostima. U nastavku neće biti tako lako direktno provjeriti je li neka funkcija $(\mathcal{F}, \mathcal{B}_T)$-izmjeriva, stoga moramo razviti jači alat kojim ćemo zadavati slučajne procese.

Lako se vidi da ako za T uzmemo $T = \{1, 2, ..., n\}$, da se definicija slučajnog procesa podudara sa definicijom slučajnoga vektora. Stoga u načelu proučavamo slučajeve kada je T beskonačan skup. Kada se u praksi uzima da je $T = \mathbb{N}, \mathbb{Z}$ ili $\mathbb{R}$, elementima $t \in T$ možemo pridružiti i značenje trenutaka u vremenu. Takve slučajne procese $\mathbb{X}$ nazivamo **vremenskim nizovima**, a $\mathbb{X}_t = \mathbb{X}(t)$ zamišljamo kao (slučajnu) vrijednost koju je $\mathbb{X}$ poprimio u trenutku $t \in T$. Nadalje, ako je $\mathbb{X}$ slučajni proces sa skupom indeksa T i $t, t_1, ..., t_n \in T$, nije teško pokazati da je $\mathbb{X}_t$ slučajna varijabla, odnosno da je $(\mathbb{X}_{t_1}, ..., \mathbb{X}_{t_n})$ slučajni vektor.

U prethodnom smo odjeljku rekli da zadati slučajni vektor $\mathbb{X} = (X_1, ..., X_n)$ po distribuciji (općenito) znači zadati njegovu funkciju distribucije $F_{\mathbb{X}} : \mathbb{R}^n \to \mathbb{R}$. Također, u napomeni 1.2.6 iskazali smo nužne i dovoljne uvjete da bi proizvoljna funkcija F bila funkcija distribucije nekog slučajnog vektora. Uz slučajni vektor $\mathbb{X}$ vežemo i vjerojatnosnu mjeru $\mathbb{P}_{\mathbb{X}}$ na izmjerivom prostoru $(\mathbb{R}^n, \mathcal{B}_n)$, danu sa

$$\mathbb{P}_{\mathbb{X}}(B) = \mathbb{P}(\mathbb{X} \in B), \quad B \in \mathcal{B}_n.$$

Iz teorije mjere (v. [4]) znamo da je (vjerojatnosnu) mjeru na $(\mathbb{R}^n, \mathcal{B}_n)$ dovoljno zadati na skupovima oblika $\langle -\infty, \mathbf{x}] = \langle -\infty, x_1] \times ... \times \langle -\infty, x_n]$, za $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. Budući imamo

$$\mathbb{P}_{\mathbb{X}}(\langle -\infty, \mathbf{x}]) = \mathbb{P}(X_1 \leq x_1, ..., X_n \leq x_n) = F_{\mathbb{X}}(\mathbf{x}), \quad \mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n, \tag{1.14}$$

time je $\mathbb{P}_{\mathbb{X}}$ zadano čim je zadano i $F_{\mathbb{X}}$. Očito vrijedi i obratno, ukoliko je zadana vjerojatnosna mjera $\mathbb{P}_{\mathbb{X}}$, tada je sa (1.14) zadana i funkcija distribucije $F_{\mathbb{X}}$. Stoga se može reći i da zadati slučajni vektor $\mathbb{X}$ po distribuciji, zapravo znači zadati vjerojatnosnu mjeru $\mathbb{P}_{\mathbb{X}}$ sa svojstvom (1.14). Pri tome ponovno ne moramo precizirati vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbb{P})$. Dapače, bez smanjenja općenitosti možemo uzeti da su $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}_n$ i $\mathbb{P} = \mathbb{P}_{\mathbb{X}}$, pri čemu je onda $\mathbb{X} : \mathbb{R}^n \to \mathbb{R}^n$ dan sa $\mathbb{X}(\omega) = \omega$, za svako $\omega \in \mathbb{R}^n$. Diskusija vrijedi i za slučajne varijable, uzimajući $n = 1$. U nastavku ovoga rada ponekad ćemo pisati i $X \sim \mathbb{P}_0$, pri čemu će se misliti na to da je distribucija slučajne varijable X zadana sa $\mathbb{P}_X = \mathbb{P}_0$. Sličnu vjerojatnosnu mjeru sada bismo htjeli uvesti za slučajne procese. To i dalje činimo po uzoru na [5], 9.4.

Definicija 1.3.4. *Neka je T proizvoljan skup indeksa i pretpostavimo da je za svako $n \in \mathbb{N}$ i za svaki neprazan konačni podskup $\{t_1, ..., t_n\}$ od T zadana vjerojatnosna konačno-dimenzionalna funkcija distribucije $F_{t_1,...,t_n} : \mathbb{R}^n \to \mathbb{R}$. Za familiju $\{F_{t_1,...,t_n}\}$ tih funkcija kažemo da zadovoljava* ***Kolmogorovljeve uvjete suglasnosti*** *ili da je* ***suglasna*** *ako su zadovoljena sljedeća dva svojstva:*

(1) Ako je $(i_1, ..., i_n)$ proizvoljna permutacija od $(1, ..., n)$, tada za svako $(x_1, ..., x_n) \in \mathbb{R}^n$ i za svaki izbor $\{t_1, ..., t_n\}$ vrijedi

$$F_{t_{i_1},...,t_{i_n}}(x_{i_1}, ..., x_{i_n}) = F_{t_1,...,t_n}(x_1, ..., x_n).$$

(2) Ako je $m < n$, tada za svako $(x_1, ..., x_m) \in \mathbb{R}^m$ i za svaki izbor $\{t_1, ..., t_n\}$ vrijedi

$$F_{t_1,...,t_m}(x_1, ..., x_m) = F_{t_1,...,t_n}(x_1, ..., x_m, \infty, ..., \infty) =$$

$$= \lim_{x_{m+1}\to\infty} ... \lim_{x_n\to\infty} F_{t_1,...,t_n}(x_1, ..., x_m, x_{m+1}, ..., x_n).$$

Teorem 1.3.5. *([5], Teorem 9.6.) Neka je T proizvoljan skup indeksa i $\{F_{t_1,...,t_n}\}$ zadana suglasna familija vjerojatnosnih konačno-dimenzionalnih funkcija distribucija. Tada postoji jedinstvena vjerojatnosna mjera $\mathbb{P}_T$ na $(\mathbb{R}^T, \mathcal{B}_T)$, koja za svaki izbor $\{t_1, ..., t_n\}$ nepraznog konačnog podskupa od T i za svaki Borelov cilinar $A = \pi^{-1}_{t_1,...,t_n}(M) \in \mathcal{F}_T$, $M \in \mathcal{B}_n$, zadovoljava*

$$\mathbb{P}_T(A) = \mathbb{P}_{t_1,...,t_n}(M), \tag{1.15}$$

gdje je $\mathbb{P}_{t_1,...,t_n}$ vjerojatnosna mjera pridružena vjerojatnosnoj konačno-dimenzionalnoj funkciji distribucije $F_{t_1,...,t_n}$.

Dakle, za zadanu suglasnu familiju $\{F_{t_1,...,t_n}\}$ vjerojatnosnih funkcija distribucija možemo promatrati vjerojatnosni prostor $(\mathbb{R}^T, \mathcal{B}_T, \mathbb{P}_T)$ i slučajni proces $\mathbb{X} : \mathbb{R}^T \to \mathbb{R}^T$ dan sa

$$\mathbb{X}(\omega) = \omega, \ \ \omega \in \mathbb{R}^T.$$

To je očito $(\mathcal{B}_T, \mathcal{B}_T)$-izmjeriva funkcija, budući je $\mathbb{X}^{-1}(A) = A$, za svako $A \in \mathcal{B}_T$. Za proizvoljne $t_1, ..., t_n \in T$ i $B_1, ..., B_n \in \mathcal{B}$, koristeći (1.15) računamo

$$\begin{aligned}
\mathbb{P}_T(\mathbb{X}_{t_1} \in B_1, ..., \mathbb{X}_{t_n} \in B_n) &= \mathbb{P}_T\left(\left\{\omega \in \mathbb{R}^T : \mathbb{X}_{t_1}(\omega) \in B_1, ..., \mathbb{X}_{t_n}(\omega) \in B_n\right\}\right) = \\
&= \mathbb{P}_T\left(\left\{\omega \in \mathbb{R}^T : \omega_{t_1} \in B_1, ..., \omega_{t_n} \in B_n\right\}\right) = \\
&= \mathbb{P}_T\left(\pi^{-1}_{t_1,...,t_n}(B_1 \times \cdots \times B_n)\right) = \\
&= \mathbb{P}_{t_1,...,t_n}(B_1 \times \cdots \times B_n).
\end{aligned}$$

Možemo reći da suglasna familija vjerojatnosnih konačno-dimenzionalnih funkcija distribucija $\{F_{t_1,...,t_n}\}$ na ovaj način zadaje slučajni proces $\mathbb{X}$, a ne samo da ga zadaje po distribuciji.

Napomena 1.3.6. Intuitivno, prvi Kolmogorovljev uvjet iz definicije 1.3.4 kaže da redoslijed u kojem promatramo slučajne varijable $\mathbb{X}_{t_1}, ..., \mathbb{X}_{t_n}$ ne utječe suštinski na distribuciju slučajnoga vektora $(\mathbb{X}_{t_1}, ..., \mathbb{X}_{t_n})$.
Drugi Kolmogorovljev uvjet osigurava da se distribucija slučajnog vektora $(\mathbb{X}_{t_1}, ..., \mathbb{X}_{t_m})$, zadana preko $F_{t_1,...,t_m}$, podudara s distribucijom koja se dobije kao marginalna distribucija prvih m komponenata slučajnoga vektora $(\mathbb{X}_{t_1}, ..., \mathbb{X}_{t_n})$, čija je distribucija zadana pomoću $F_{t_1,...,t_n}$.

Primjer 1.3.7. Neka su $T = \mathbb{N}$ te $F_0 : \mathbb{R} \to \mathbb{R}$ neka vjerojatnosna funkcija distribucije. Neka je za svaki izbor $n \in \mathbb{N}$ te $t_1, ..., t_n \in T$ zadana vjerojatnosna konačno-dimenzionalna funkcija distribucije

$$F_{t_1,...,t_n}(x_1, ..., x_n) = \prod_{i=1}^{n} F_0(x_i), \quad (x_1, ..., x_n) \in \mathbb{R}^n.$$

Nije se teško uvjeriti da je tako zadana $F_{t_1,...,t_n}$ zaista vjerojatnosna konačno-dimenzionalna funkcija distribucije, čim je F_0 vjerojatnosna funkcija distribucije. Provjerimo da je tada

$\{F_{t_1,...,t_n}\}$ suglasna familija. Neka su $n \in \mathbb{N}$ te $\{t_1, ..., t_n\} \subset T$. Neka je $(i_1, ..., i_n)$ proizvoljna permutacija od $(1, ..., n)$. Očito je

$$F_{t_{i_1},...,t_{i_n}}(x_{i_1}, ..., x_{i_n}) = \prod_{j=1}^{n} F_0(x_{i_j}) = \prod_{i=1}^{n} F_0(x_i) = F_{t_1,...,t_n}(x_1, ..., x_n),$$

za svako $(x_1, ..., x_n) \in \mathbb{R}^n$. Stoga je zadovoljen prvi Kolmogorovljev uvjet suglasnosti iz definicije 1.3.4. Ako su pak $m < n$ te $(x_1, ..., x_m) \in \mathbb{R}^m$, tada je

$$F_{t_1,...,t_n}(x_1, ..., x_m, \infty, ..., \infty) = \prod_{i=1}^{m} F_0(x_i) \cdot \prod_{i=m+1}^{n} F_0(\infty) = \prod_{i=1}^{m} F_0(x_i) \cdot 1 = F_{t_1,...,t_m}(x_1, ..., x_m),$$

gdje je $F_0(\infty) = \lim_{x\to\infty} F_0(x) = 1$, budući je F_0 vjerojatnosna funkcija distribucije. Dakle, zadovoljen je i drugi Kolmogorovljev uvjet suglasnosti iz definicije 1.3.4. Prema teoremu 1.3.5, familija $\{F_{t_1,...,t_n}\}$ zadaje neki slučajni proces $\mathbb{X}$. Štoviše, za taj slučajni proces $\mathbb{X}$ te za svaki izbor n različitih $t_1, ..., t_n \in T$, slučajni je vektor $(\mathbb{X}_{t_1}, ...\mathbb{X}_{t_n})$ sastavljen od n nezavisnih i jednako distribuiranih slučajnih varijabli s funkcijom distribucije upravo F_0.

Poglavlje 2

Bayesovska statistika

U ovom poglavlju uvodimo osnovne pojmove tzv. *parametarske bayesovske statistike*. To činimo po uzoru na [2]. Teorija će biti potkrijepljena primjerima. Način zaključivanja koji ćemo izložiti poslužit će nam kao motivacija za poglavlje 3 i uvod u tzv. *neparametarsku bayesovsku statistiku*.

2.1 Bayesova formula. Motivacija

Neka je dan vjerojatnosni prostor $(\Omega, \mathcal{F}, \mathbb{P})$. Pretpostavimo da su $A, B \in \mathcal{F}$ događaji te da vrijedi $\mathbb{P}(B) > 0$. Uvjetnu vjerojatnost događaja A, uz dano B definiramo sa

$$\mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Uvjetnu vjerojatnost $\mathbb{P}(A \,|\, B)$ interpretiramo kao vjerojatnost događaja A, ukoliko nam je poznato da se dogodilo B. Nije teško pokazati da je tako uvedena $\mathbb{P}_B : \mathcal{F} \to \mathbb{R}$, dana sa $\mathbb{P}_B(A) = \mathbb{P}(A \,|\, B)$, $A \in \mathcal{F}$, ponovno vjerojatnost na $(\Omega, \mathcal{F})$. Specijalno, vrijedi $\mathbb{P}_B(\Omega) = 1$.

Sada možemo vrlo jednostavno pokazati *Bayesovu formulu*

$$\mathbb{P}(A \,|\, B) = \mathbb{P}(B \,|\, A) \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{2.1}$$

Zaista, ako su $\mathbb{P}(A)$, $\mathbb{P}(B) > 0$, imamo

$$\mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \mathbb{P}(B \,|\, A) \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Prije nego krenemo uvoditi pojmove vezane uz bayesovsku statistiku, pogledajmo jedan primjer u kojem ćemo iskoristiti (2.1). Primjer je preuzet iz [2], 1.4.

Primjer 2.1.1. Iz genetike znamo da se u čovjekovom genomu javljaju dvije vrste spolnih kromosoma koje označavamo slovima X i Y. Ženske osobe imaju dva istovrsna X kromosoma, dok muške osobe imaju po jedan X i Y kromosom. Prilikom oplodnje, u načelu, začeto dijete nasljeđuje po jedan nasumični spolni kromosom od svakog roditelja.
Hemofilija je genetska bolest koja se prenosi X kromosomom, tj. gen za hemofiliju se može nalaziti na X kromosomu. Ženske osobe s tim genom na samo jednom X kromosomu u pravilu nemaju nikakvih simptoma i nisu svjesne da problematični gen mogu prenijeti na svoje potomke. S druge strane, muške osobe koje su od nekog roditelja naslijedile gen za hemofiliju gotovo uvijek pokazuju znakove bolesti. Stoga se za hemofiliju kaže da je muška bolest koja se prenosi po majčinoj liniji. Moguća je i situacija u kojoj se žensko dijete rađa s dva gena za hemofiliju (po jednim na svakom X kromosomu). Takve osobe imaju vrlo teške simptome bolesti. Međutim, ti su slučajevi vrlo rijetki, a tim više što se muškarci s hemofilijom uglavnom odlučuju ne imati djecu.
Promatramo ženu čiji brat ima dijagnosticiranu hemofiliju, koju je sigurno naslijedio od njihove zajedničke majke. Pitamo se kolika je vjerojatnost da je i žena naslijedila X kromosom s genom za hemofiliju. Neka je

$$Y = \begin{cases} 0, & \text{ako žena } \textbf{nije} \text{ naslijedila gen za hemofiliju,} \\ 1, & \text{ako žena } \textbf{jest} \text{ naslijedila gen za hemofiliju.} \end{cases}$$

Bez dodatnih saznanja, najbolje što možemo reći jest da je prilikom oplodnje postojala jednaka vjerojatnost da žena od majke naslijedi X kromosom s genom za hemofiliju i X kromosom bez njega. Dakle,

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{1}{2}.$$

Recimo da je žena potom rodila sina koji ne pokazuje znakove bolesti. Pišemo $X_1 = 0$, tj. neka X_1, slično kao Y, označava je li sin od majke naslijedio gen za hemofiliju ($X_1 = 1$) ili nije ($X_1 = 0$). (Pri tome je jasno da muško dijete nije naslijedilo hemofiliju od oca). U skladu s gore opisanim osnovama genetike, imamo

$$\mathbb{P}(X_1 = 0 \,|\, Y = 1) = \frac{1}{2},$$

odnosno

$$\mathbb{P}(X_1 = 0 \,|\, Y = 0) = 1.$$

Primijenjujući (2.1) dobivamo:

$$\mathbb{P}(Y = 1 \,|\, X_1 = 0) = \mathbb{P}(X_1 = 0 \,|\, Y = 1) \cdot \frac{\mathbb{P}(Y = 1)}{\mathbb{P}(X_1 = 0)} = \frac{1}{2} \cdot \frac{½}{\mathbb{P}(X_1 = 0)} = \frac{1}{4} \cdot \frac{1}{\mathbb{P}(X_1 = 0)},$$

odnosno

$$\mathbb{P}(Y = 0 \,|\, X_1 = 0) = \mathbb{P}(X_1 = 0 \,|\, Y = 0) \cdot \frac{\mathbb{P}(Y = 0)}{\mathbb{P}(X_1 = 0)} = 1 \cdot \frac{1/2}{\mathbb{P}(X_1 = 0)} = \frac{1}{2} \cdot \frac{1}{\mathbb{P}(X_1 = 0)}.$$

Na prvu nije jasno kako bismo izračunali $\mathbb{P}(X_1 = 0)$, međutim, dovoljno je sjetiti se da vrijedi

$$\mathbb{P}(Y = 1 \,|\, X_1 = 0) + \mathbb{P}(Y = 0 \,|\, X_1 = 0) = 1,$$

iz čega onda slijedi

$$\mathbb{P}(Y = 1 \,|\, X_1 = 0) = \frac{1}{3},$$

odnosno

$$\mathbb{P}(Y = 0 \,|\, X_1 = 0) = \frac{2}{3}.$$

Drugim riječima, nakon što je žena rodila zdravoga sina, skloniji smo vjerovati da žena nema gen za hemofiliju.

Pretpostavimo da je nakon prvoga zdravog sina, žena rodila još jednoga sina bez hemofilije (analogno kao prije, pišemo $X_2 = 0$). Budući se oplodnja kod drugoga sina odvijala nezavisno od oplodnje kod prvoga, slično kao i ranije imamo

$$\mathbb{P}(X_1 = 0, X_2 = 0 \,|\, Y = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

te

$$\mathbb{P}(X_1 = 0, X_2 = 0 \,|\, Y = 0) = 1 \cdot 1 = 1.$$

Koristeći (2.1) dobivamo

$$\begin{aligned}\mathbb{P}(Y = 1 \,|\, X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 0 \,|\, Y = 1) \cdot \frac{\mathbb{P}(Y = 1)}{\mathbb{P}(X_1 = 0, X_2 = 0)} = \\ = \frac{1}{4} \cdot \frac{1/2}{\mathbb{P}(X_1 = 0, X_2 = 0)} = \frac{1}{8} \cdot \frac{1}{\mathbb{P}(X_1 = 0, X_2 = 0)},\end{aligned}$$

odnosno

$$\begin{aligned}\mathbb{P}(Y = 0 \,|\, X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 0 \,|\, Y = 0) \cdot \frac{\mathbb{P}(Y = 0)}{\mathbb{P}(X_1 = 0, X_2 = 0)} = \\ = 1 \cdot \frac{1/2}{\mathbb{P}(X_1 = 0, X_2 = 0)} = \frac{1}{2} \cdot \frac{1}{\mathbb{P}(X_1 = 0, X_2 = 0)}.\end{aligned}$$

Primjenom istoga *trika* kao i ranije slijedi

$$\mathbb{P}(Y = 1 \,|\, X_1 = 0, X_2 = 0) = \frac{1}{5},$$

odnosno

$$\mathbb{P}(Y = 0 \,|\, X_1 = 0, X_2 = 0) = \frac{4}{5}.$$

Intuitivno, rođenjem svakog novog zdravog muškog djeteta, sve smo uvjereniji da žena nema kromosom s genom za hemofiliju.

2.2 Apriorna i aposteriorna gustoća

U primjeru 2.1.1 htjeli smo odgovoriti na pitanje *koliko je vjerojatno* da promatrana žena ima gen za hemofiliju. Budući da žena ili ima spomenuti gen ili ga nema, upitno je što bi uopće značila izjava poput *"vjerojatnost da žena ima gen za hemofiliju je 20%."* Iako je u sadašnjosti zbilja istina da žena ili ima taj gen ili ga nema, svjesni smo da je prilikom oplodnje došlo do realizacije svojevrsnoga slučajnog pokusa, odnosno da je postojala vjerojatnost od 50% da se problematični gen naslijedi i vjerojatnost od 50% da se ne naslijedi. Frekvencionistički govoreći, kada bismo promatrali velik broj žena čije majke potvrđeno imaju gen za hemofiliju, očekivali bismo da je oko polovica tih žena isti gen i naslijedila. Stoga smo krenuli s tim modelom:

$$Y \sim \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}, \text{ tj. } Y \sim B(1/2)\,.$$

Nakon toga smo uveli slučajne varijable X_1, X_2 i uočili da kada bismo poznavali stvarnu vrijednost od Y, znali bismo distribucije od X_1, odnosno X_2. Ne možemo direktno opažati veličinu Y, ali možemo X_1 i X_2. Stoga smo iskoristili Bayesovu formulu (2.1) kako bismo, uvjetno na konkretne vrijednosti od X_1 i X_2 (kojima raspolažemo), rekli nešto novo o distribuciji slučajne varijable Y.

Pojam vjerojatnosti u bayesovskoj statistici predstavlja subjektivnu mjeru uvjerenosti u neki događaj. Ta se uvjerenost temelji na određenom polaznom modelu i opaženim podacima.

Uočimo da nije određeno koju distribuciju slijede slučajne varijable X_1 i X_2. Kao što je već bilo rečeno, distribucije od X_1, X_2 bile bi poznate tek nakon što bismo znali da je nastupio događaj $\{Y = 0\}$ ili $\{Y = 1\}$. Dakle, zapravo za svaku moguću vrijednost $y \in \mathbb{R}$, koju poprima slučajna varijabla Y, promatramo zasebne vjerojatnosne distribucije

$$X_i^{(Y=0)} \sim B(0), \quad X_i^{(Y=1)} \sim B(1/2)\,, \quad i = 1, 2.$$

Budući su tako uzete slučajne varijable $X_i^{(Y=0)}, X_i^{(Y=1)}$, $i = 1, 2$ diskretne, imamo

$$f_{X_i^{(Y=0)}}(x) = \mathbb{P}(X_i = x \,|\, Y = 0) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases} \qquad x \in \mathbb{R},$$

odnosno

$$f_{X_i^{(Y=1)}}(x) = \mathbb{P}(X_i = x \,|\, Y = 1) = \begin{cases} \frac{1}{2}, & x = 0, 1 \\ 0, & x \neq 0, 1 \end{cases} \qquad x \in \mathbb{R},$$

za $i = 1, 2$. Drugi način na koji bismo to mogli zapisati je

$$f_{X_i \,|\, Y}(x \,|\, y) = f_{X_i^{(Y=y)}}(x) = \begin{cases} 1 - \frac{1}{2}y, & x = 0 \\ \frac{1}{2}y, & x = 1 \\ 0, & x \neq 0, 1 \end{cases} \qquad x \in \mathbb{R},\ y = 0, 1,$$

za $i = 1, 2$. Ovakvu funkciju gustoće $f_{X_i \,|\, Y}$, koja je ujedno i funkcija argumenta y, nazivat ćemo *funkcijom gustoće slučajne varijable X_i uvjetno na Y*. Formalnu definiciju ovog pojma iskazujemo za sve moguće slučajeve koje ćemo koristiti u nastavku. Do kraja odjeljka uzimamo da su $n, m \in \mathbb{N}$.

Definicija 2.2.1. *Neka su na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ dani slučajni vektori $\mathbb{X} = (X_1, ..., X_n)$ i $\mathbb{Y} = (Y_1, ..., Y_m)$ (diskretni ili neprekidni). Neka su $f_{\mathbb{X}}$ i $f_{\mathbb{Y}}$ pripadne funkcije gustoća tih vektora, a $f_{\mathbb{X},\mathbb{Y}}$ gustoća slučajnog vektora $(\mathbb{X}, \mathbb{Y})$ (v. odjeljak 1.2). Označimo sa*

$$C_{\mathbb{Y}} = \left\{ \boldsymbol{y} \in \mathbb{R}^m : f_{\mathbb{Y}}(\boldsymbol{y}) > 0 \right\}.$$

Funkciju $f_{\mathbb{X}|\mathbb{Y}} : \mathbb{R}^n \times C_{\mathbb{Y}} \to \mathbb{R}$ danu sa

$$f_{\mathbb{X}|\mathbb{Y}}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \frac{f_{\mathbb{X},\mathbb{Y}}(\boldsymbol{x}, \boldsymbol{y})}{f_{\mathbb{Y}}(\boldsymbol{y})}, \quad \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{y} \in C_{\mathbb{Y}}$$

nazivamo ***funkcijom gustoće slučajnog vektora $\mathbb{X}$ uvjetno na $\mathbb{Y}$***.

Uočimo da ova definicija ima smisla i za $n = 1$, odnosno $m = 1$. Tada govorimo o **funkciji gustoće slučajne varijable uvjetno na slučajni vektor** ($n = 1$, $m > 1$), **funkciji gustoće slučajnog vektora uvjetno na slučajnu varijablu** ($n > 1$, $m = 1$), odnosno o **funkciji gustoće slučajne varijable uvjetno na slučajnu varijablu** ($n = 1$, $m = 1$). Analogno definiramo i

$$f_{\mathbb{Y}|\mathbb{X}}(\mathbf{y} \,|\, \mathbf{x}) = \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})}, \quad \mathbf{y} \in \mathbb{R}^m,\ \mathbf{x} \in C_{\mathbb{X}},$$

gdje je $C_{\mathbb{X}} = \{\, \mathbf{x} \in \mathbb{R}^n : f_{\mathbb{X}}(\mathbf{x}) > 0 \,\}$.

Napomena 2.2.2. U nastavku rada pretpostavljamo da su funkcije gustoća neprekidnih slučajnih vektora neprekidne. Također, ukoliko je $\mathbb{X}$ diskretan, a $\mathbb{Y}$ neprekidan slučajni vektor, pretpostavljamo da je $\mathbf{y} \mapsto f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})$ neprekidno preslikavanje za svako fiksno $\mathbf{x} \in \mathbb{R}^n$. Ukoliko je $\mathbb{X}$ neprekidan, a $\mathbb{Y}$ diskretan slučajni vektor, pretpostavljamo da je $\mathbf{x} \mapsto f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})$ neprekidno preslikavanje za svako fiksno $\mathbf{y} \in \mathbb{R}^m$.

Propozicija 2.2.3. *Neka su $\mathbb{X}$ i $\mathbb{Y}$ slučajni vektori te $f_{\mathbb{X}}, f_{\mathbb{Y}}$ i $f_{\mathbb{X},\mathbb{Y}}$ funkcije kao u definiciji 2.2.1. Dodatno pretpostavljamo tvrdnju napomene 2.2.2. Tada je preslikavanje $\boldsymbol{x} \mapsto f_{\mathbb{X}|\mathbb{Y}}(\boldsymbol{x} \,|\, \boldsymbol{y})$ funkcija gustoće nekog n-dimenzionalnog slučajnog vektora, za svako $\boldsymbol{y} \in \mathbb{R}^m$ takvo da je $f_{\mathbb{Y}}(\boldsymbol{y}) > 0$. (Analogno vrijedi i za $f_{\mathbb{Y}|\mathbb{X}}$).*

Dokaz. Moramo provjeriti tvrdnju za četiri različita slučaja.

(I) Neka su $\mathbb{X}$ i $\mathbb{Y}$ diskretni slučajni vektori. (Tada je to i slučajni vektor $(\mathbb{X}, \mathbb{Y})$). Za proizvoljno $\mathbf{y} \in \mathbb{R}^m$ takvo da je $f_{\mathbb{Y}}(\mathbf{y}) > 0$ imamo

$$\sum_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y}) = \sum_{\mathbf{x}\in\mathbb{R}^n} \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbb{Y}}(\mathbf{y})} = \frac{\mathbb{P}(\mathbb{X} \in \mathbb{R}^n, \mathbb{Y} = \mathbf{y})}{\mathbb{P}(\mathbb{Y} = \mathbf{y})} = \frac{\mathbb{P}(\mathbb{Y} = \mathbf{y})}{\mathbb{P}(\mathbb{Y} = \mathbf{y})} = 1,$$

pri čemu je $f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y}) \neq 0$ za najviše prebrojivo mnogo $\mathbf{x} \in \mathbb{R}^n$ (jer je već $f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y}) \neq 0$ za najviše prebrojivo mnogo $\mathbf{x} \in \mathbb{R}^n$). Kako je $f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y}) \geq 0$ za svako $\mathbf{x} \in \mathbb{R}^n$, imamo da je $\mathbf{x} \mapsto f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y})$ funkcija gustoće nekog diskretnog slučajnog vektora.

(II) Neka su $\mathbb{X}$ i $\mathbb{Y}$ neprekidni slučajni vektori. (Po pretpostavci je to tada i $(\mathbb{X}, \mathbb{Y})$). Neka je $B_m \in \mathcal{B}_m$ proizvoljno. Tada imamo

$$\int_{B_m} \int_{\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y}) d\lambda^n(\mathbf{x}) d\lambda^m(\mathbf{y}) = \mathbb{P}(\mathbb{X} \in \mathbb{R}^n, \mathbb{Y} \in B_m) = \mathbb{P}(\mathbb{Y} \in B_m) = \int_{B_m} f_{\mathbb{Y}}(\mathbf{y}) d\lambda^m(\mathbf{y}).$$

Budući su $\mathbf{y} \mapsto f_{\mathbb{Y}}(\mathbf{y})$ i $\mathbf{y} \mapsto f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})$ po pretpostavci neprekidne funkcije (za svako $\mathbf{x} \in \mathbb{R}^n$), a samim time i $\mathbf{y} \mapsto \int_{\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y}) \, d\lambda^n(\mathbf{x})$, zbog proizvoljnosti od $B_m \in \mathcal{B}_m$, prema [4] slijedi da je

$$\int_{\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y}) \, d\lambda^n(\mathbf{x}) = f_{\mathbb{Y}}(\mathbf{y})$$

za svako $\mathbf{y} \in \mathbb{R}^m$. Od tuda za proizvoljno $\mathbf{y} \in \mathbb{R}^m$, takvo da je $f_{\mathbb{Y}}(\mathbf{y}) > 0$, dobivamo

$$1 = \frac{\int_{\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y}) \, d\lambda^n(\mathbf{x})}{f_{\mathbb{Y}}(\mathbf{y})} = \int_{\mathbb{R}^n} f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y}) \, d\lambda^n(\mathbf{x}).$$

Kako je funkcija $\mathbf{x} \mapsto f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y})$ očito nenegativna i Borelova, to je ona funkcija gustoće nekog neprekidnog slučajnog vektora.

(III) Neka je $\mathbb{X}$ diskretan, a $\mathbb{Y}$ neprekidan slučajni vektor. Neka je $B_m \in \mathcal{B}_m$ proizvoljno. Tada imamo

$$\int_{B_m} \sum_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})\, d\lambda^m(\mathbf{y}) = \mathbb{P}(\mathbb{X} \in \mathbb{R}^n, \mathbb{Y} \in B_m) = \mathbb{P}(\mathbb{Y} \in B_m) = \int_{B_m} f_{\mathbb{Y}}(\mathbf{y})\, d\lambda^m(\mathbf{y}).$$

Kao i u prethodnom slučaju, zbog neprekidnosti funkcija $\mathbf{y} \mapsto \sum_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})$ i $\mathbf{y} \mapsto f_{\mathbb{Y}}(\mathbf{y})$ te zbog proizvoljnosti od $B_m \in \mathcal{B}_m$, prema [4] slijedi da je

$$\sum_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbb{Y}}(\mathbf{y})$$

za svako $\mathbf{y} \in \mathbb{R}^m$. Slično, to da je i u ovom slučaju $f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y}) \neq 0$ za najviše prebrojivo mnogo $\mathbf{x} \in \mathbb{R}^n$ slijedi iz (1.12). Naime, zbog proizvoljnosti od $B_m \in \mathcal{B}_m$ u (1.12) te neprekidnosti i nenegativnosti preslikavanja $\mathbf{y} \mapsto f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})$, imamo da je $f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y}) = 0$, $\forall \mathbf{y} \in \mathbb{R}^m$, čim je $f_{\mathbb{X}}(\mathbf{x}) = 0$. Sada za proizvoljno $\mathbf{y} \in \mathbb{R}^m$, takvo da je $f_{\mathbb{Y}}(\mathbf{y}) > 0$, dobivamo

$$1 = \frac{\sum_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbb{Y}}(\mathbf{y})} = \sum_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y}),$$

pri čemu je $f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y}) \neq 0$ za najviše prebrojivo mnogo $\mathbf{x} \in \mathbb{R}^n$ (jer je već $f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y}) \neq 0$ za najviše prebrojivo mnogo $\mathbf{x} \in \mathbb{R}^n$). Kako je $f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y}) \geq 0$ za svako $\mathbf{x} \in \mathbb{R}^n$, imamo da je $\mathbf{x} \mapsto f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y})$ funkcija gustoće nekog diskretnog slučajnog vektora.

(IV) Neka je $\mathbb{X}$ neprekidan, a $\mathbb{Y}$ diskretan slučajni vektor. Za proizvoljno $\mathbf{y} \in \mathbb{R}^m$ takvo da je $f_{\mathbb{Y}}(\mathbf{y}) > 0$, imamo

$$\int_{\mathbb{R}^n} f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y})\, d\lambda^n(\mathbf{x}) = \int_{\mathbb{R}^n} \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbb{Y}}(\mathbf{y})}\, d\lambda^n(\mathbf{x}) = \frac{\mathbb{P}(\mathbb{X} \in \mathbb{R}^n, \mathbb{Y} = \mathbf{y})}{\mathbb{P}(\mathbb{Y} = \mathbf{y})} = \frac{\mathbb{P}(\mathbb{Y} = \mathbf{y})}{\mathbb{P}(\mathbb{Y} = \mathbf{y})} = 1.$$

Kako je funkcija $\mathbf{x} \mapsto f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y})$ očito nenegativna i Borelova, to je ona funkcija gustoće nekog neprekidnog slučajnog vektora.

□

Teorem 2.2.4. ***(Bayesov teorem)*** *Neka su $\mathbb{X}$ i $\mathbb{Y}$ slučajni vektori te $f_{\mathbb{X}}, f_{\mathbb{Y}}$ i $f_{\mathbb{X},\mathbb{Y}}$ funkcije kao u definiciji 2.2.1. Tada za svako $\boldsymbol{x} \in \mathbb{R}^n$ i $\boldsymbol{y} \in \mathbb{R}^m$ takve da je $f_{\mathbb{X}}(\boldsymbol{x}) > 0$ i $f_{\mathbb{Y}}(\boldsymbol{y}) > 0$ vrijedi*

$$f_{\mathbb{Y}|\mathbb{X}}(\boldsymbol{y}\,|\,\boldsymbol{x}) = f_{\mathbb{X}|\mathbb{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) \cdot \frac{f_{\mathbb{Y}}(\boldsymbol{y})}{f_{\mathbb{X}}(\boldsymbol{x})}. \tag{2.2}$$

Dokaz.

$$f_{\mathbb{Y}|\mathbb{X}}(\mathbf{y}\,|\,\mathbf{x}) = \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})} = \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})} \cdot \frac{f_{\mathbb{Y}}(\mathbf{y})}{f_{\mathbb{Y}}(\mathbf{y})} = \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbb{Y}}(\mathbf{y})} \cdot \frac{f_{\mathbb{Y}}(\mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})} = f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}\,|\,\mathbf{y}) \cdot \frac{f_{\mathbb{Y}}(\mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})}.$$

□

U statistici često rješavamo sljedeći problem. Promatramo *(populacijsko) statističko obilježje*, odnosno slučajnu varijablu X. Ta slučajna varijabla ima svoju funkciju gustoće f_X koja ovisi o nekom parametru $p \in \mathbb{R}$ (vidi primjere 1.1.6 i 1.1.9). Taj parametar je, međutim, nepoznat. Nezavisne slučajne varijable $X_1, ..., X_n$ ($n \in \mathbb{N}$), koje su sve distribuirane jednako kao X, nazivamo **slučajnim uzorkom**. Konkretne realizacije $x_1, ..., x_n \in \mathbb{R}$ slučajnoga uzorka nazivamo **uzorkom**. Na temelju uzorka želimo procijeniti nepoznati parametar p. Frekvencionistička statistika pretpostavlja da je parametar p nepoznat, ali fiksan, tj. da se radi o realnom broju. S druge strane, u bayesovskoj statistici uzimamo da je parametar p nepoznat i slučajan, odnosno da se radi o slučajnoj varijabli. Iako je slučajne varijable uobičajeno označavati velikim slovima latinice (npr. X, Y, Z), parametre najčešće označavamo malim slovima latinice ili grčkoga alfabeta (npr. p, a, b, μ, θ). Stoga ćemo i u ovome kontekstu parametre (koji su ujedno i slučajne varijable) označavati malim slovima. Dakle, nama je zapravo dana $f_{X|p}$ funkcija gustoće slučajne varijable X uvjetno na p. Neka je uz to dana i f_p funkcija gustoće parametra p. Ukoliko za slučajne varijable $X_1, ..., X_n$ vrijedi

$$f_{X_1|p} = ... = f_{X_n|p} ,$$

reći ćemo da su $X_1, ..., X_n$ **jednako distribuirane uvjetno na** p. Konkretno, uzimamo da je

$$f_{X_i|p} = f_{X|p}, \quad i = 1, ..., n.$$

Također, ako za svako $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ i za svako $p_0 \in \mathbb{R}$, takvo da je $f_p(p_0) > 0$, vrijedi

$$f_{\mathbb{X}|p}(\mathbf{x} \,|\, p_0) = \prod_{i=1}^{n} f_{X_i|p}(x_i \,|\, p_0),$$

gdje je $\mathbb{X} = (X_1, ..., X_n)$, po uzoru na formulu (1.13), reći ćemo da su $X_1, ..., X_n$ **nezavisne uvjetno na** p. Mi ćemo uzimati da su $X_1, ..., X_n$ jednako distribuirane uvjetno na p i nezavisne uvjetno na p, odnosno kraće kažemo da su **nezavisne i jednako distribuirane uvjetno na** p. Uočimo da to ne znači da su $X_1, ..., X_n$ same po sebi nezavisne niti da su jednako distribuirane, tj. da ne vrijedi nužno

$$f_{\mathbb{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall \mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n,$$

$$f_{X_1} = ... = f_{X_n}.$$

Kako smo orijentirani na funkcije $f_{\mathbb{X}|p}$ i $f_{X_i|p}$ (umjesto na $f_{\mathbb{X}}$ i f_{X_i}), $(i = 1, ..., n)$, niz slučajnih varijabli $X_1, ..., X_n$ ipak možemo zvati slučajnim uzorkom u bayesovskom smislu. Štoviše, ako kažemo da $X_1, ..., X_n$ slijede neku distribuciju s parametrom p, pripadno ćemo misliti na uvjetne funkcije gustoće $f_{X_i|p}$, a ne na f_{X_i}.

Budući smo pretpostavili postojanje funkcije $f_{\mathbb{X}|p}$, time smo, uz gustoću f_p, prešutno pretpostavili i postojanje gustoća $f_{\mathbb{X}}$ i $f_{\mathbb{X},p}$ (v. definiciju 2.2.1). Dodatno, ukoliko $\mathbb{X}$ uvjetno na p slijedi neku diskretnu distribuciju, pretpostavljat ćemo da je i $\mathbb{X}$ diskretan slučajni vektor s gustoćom $f_{\mathbb{X}}$. Analogno ukoliko $\mathbb{X}$ uvjetno na p slijedi neku neprekidnu distribuciju. Pokazat će se da nam to što je $f_{\mathbb{X}}$ nepoznato, neće stvarati problem.

Funkciju gustoće f_p parametra p nazivamo **apriornom gustoćom** parametra p. Njome je zadana tzv. **apriorna distribucija (razdioba)** od p (eng. *prior distribution*), koju uzimamo kao važeći početni model. Nakon toga opažamo konkretne vrijednosti $x_1, ..., x_n \in \mathbb{R}$ koje redom poprimaju slučajne varijable $X_1, ..., X_n$. Koristeći formulu (2.2) možemo predložiti novu distribuciju parametra p, danu sa $f_{p|\mathbb{X}}(\cdot\,|\,\mathbf{x})$, tj. funkcijom gustoće parametra p uvjetno na $\mathbb{X} = (X_1, ..., X_n)$, pri čemu je $\mathbf{x} = (x_1, ..., x_n)$ poznato i fiksno. Takvu distribuciju nazivamo **aposteriornom distribucijom (razdiobom)** parametra p (eng. *posterior distribution*). Pri tome pretpostavljamo da je $f_{\mathbb{X}}(\mathbf{x}) > 0$, što ima smisla, budući su se vrijednosti $x_1, ..., x_n$ pojavile kao realizacije od $X_1, ..., X_n$. U tom se slučaju formula (2.2) svodi na

$$f_{p|\mathbb{X}}(p_0\,|\,\mathbf{x}) = f_{\mathbb{X}|p}(\mathbf{x}\,|\,p_0) \cdot \frac{f_p(p_0)}{f_{\mathbb{X}}(\mathbf{x})} = \left(\prod_{i=1}^{n} f_{X_i|p}(x_i\,|\,p_0)\right) \cdot \frac{f_p(p_0)}{f_{\mathbb{X}}(\mathbf{x})}, \quad p_0 \in \mathbb{R},\ f_p(p_0) > 0. \tag{2.3}$$

Izraz

$$f_{\mathbb{X}|p}(\mathbf{x}\,|\,p_0) = \prod_{i=1}^{n} f_{X_i|p}(x_i\,|\,p_0)$$

nam je poznat iz frekvencionističke statistike te ga, kao i tamo, nazivamo **vjerodostojnost** (eng. *likelihood*). Kako je $f_{\mathbb{X}}(\mathbf{x})$ fiksno (jer je $\mathbf{x} \in \mathbb{R}^n$ fiksno), označimo

$$\frac{1}{f_{\mathbb{X}}(\mathbf{x})} =: c \in \mathbb{R} \setminus \{0\}.$$

Sada se formula (2.3) pojednostavljuje na

$$f_{p|\mathbb{X}}(p_0\,|\,\mathbf{x}) = c \cdot f_p(p_0) \cdot \prod_{i=1}^{n} f_{X_i|p}(x_i\,|\,p_0), \quad p_0 \in \mathbb{R},\ f_p(p_0) > 0. \tag{2.4}$$

Konstantu c nije teško izračuati uzmemo li u obzir da je $p_0 \mapsto f_{p|\mathbb{X}}(p_0\,|\,\mathbf{x})$ funkcija distribucije (za $p_0 \in \mathbb{R}$ takvo da je $f_p(p_0) = 0$ možemo staviti $f_{p|\mathbb{X}}(p_0\,|\,\mathbf{x}) = 0$) pa mora vrijediti

$$\int_{\mathbb{R}} f_{p|\mathbb{X}}(p_0\,|\,\mathbf{x})\ d\lambda(p_0) = 1,$$

ukoliko je p neprekidna, odnosno

$$\sum_{p_0 \in \mathbb{R}} f_{p|\mathbb{X}}(p_0\,|\,\mathbf{x}) = 1,$$

ukoliko je p diskretna slučajna varijabla. Konačno, formulu (2.4) možemo zapisati kao

$$f_{p|\mathbb{X}}(p_0 \,|\, \mathbf{x}) \propto f_p(p_0) \cdot \prod_{i=1}^{n} f_{X_i|p}(x_i \,|\, p_0), \quad p_0 \in \mathbb{R},\ f_p(p_0) > 0. \tag{2.5}$$

Naime, ukoliko su dane proizvoljne funkcije $f, g : \mathbb{R} \to \mathbb{R}$ te postoji $c \in \mathbb{R} \setminus \{0\}$ takvo da je

$$f(x) = c \cdot g(x), \quad \forall x \in \mathbb{R}\,,$$

kažemo da su f i g *proporcionalne* i pišemo $f(x) \propto g(x)$. Koristeći terminologiju bayesovske statistike, možemo reći da je osnovna shema bayesovskog zaključivanja dana izrazom

$$\textit{posterior} \propto \textit{prior} \cdot \textit{likelihood}.$$

Rezultat ovakve analize nije brojčana procjena parametra p, kao što je to slučaj u frekvencionističkoj statistici, već čitava distribucija. Ukoliko bismo ipak htjeli navesti neki konkretan broj, kao procjenitelj ima smisla uzeti očekivanje aposteriorne distribucije. Taj se procjenitelj naziva **bayesovski procjenitelj**.

Napomena 2.2.5. Ukoliko su $\mathbb{X}, \mathbb{Y}$ diskretni slučajni vektori, uočimo da formula (2.2) slijedi direktno iz (2.1). Naime

$$f_{\mathbb{Y}|\mathbb{X}}(\mathbf{y} \,|\, \mathbf{x}) = \frac{f_{\mathbb{X},\mathbb{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})} = \frac{\mathbb{P}(\mathbb{X} = \mathbf{x}, \mathbb{Y} = \mathbf{y})}{\mathbb{P}(\mathbb{X} = \mathbf{x})} = \mathbb{P}(\mathbb{Y} = \mathbf{y} \,|\, \mathbb{X} = \mathbf{x}) =$$

$$= \mathbb{P}(\mathbb{X} = \mathbf{x} \,|\, \mathbb{Y} = \mathbf{y}) \cdot \frac{\mathbb{P}(\mathbb{Y} = \mathbf{y})}{\mathbb{P}(\mathbb{X} = \mathbf{x})} = f_{\mathbb{X}|\mathbb{Y}}(\mathbf{x} \,|\, \mathbf{y}) \cdot \frac{f_{\mathbb{Y}}(\mathbf{y})}{f_{\mathbb{X}}(\mathbf{x})}.$$

Ipak, parametar čiju aposteriornu distribuciju želimo procijeniti je u praksi rijetko diskretan. U primjeru 2.1.1 pokazano je na koji način možemo pristupiti tom problemu. On je poslužio više kao motivacija. U nastavku ćemo isključivo uzimati da je nepoznati parametar neprekidna slučajna varijabla. Treba napomenuti da u tom slučaju formula (2.2) ipak donosi nešto suštinski novo, u odnosu na formulu (2.1). Naime, ukoliko je Y neprekidna slučajna varijabla, za svako $y \in \mathbb{R}$ imamo $\mathbb{P}(Y = y) = 0$ pa $\mathbb{P}(\mathbb{X} = \mathbf{x} \,|\, Y = y)$ uopće nije definirano. Stoga je bilo nužno na drugačiji način uvesti pojmove uvjetnih funkcija gustoća i pokazati formulu (2.2), koju također nazivamo **Bayesovom formulom**.

2.3 Primjeri

Pogledajmo dva primjera u kojima zaključujemo aposteriornu distribuciju nepoznatoga parametra. U prvom je primjeru parametar neprekidna, a elementi slučajnog uzorka su diskretne slučajne varijable.

Primjer 2.3.1. U nekom se gradu održavaju izbori za gradonačelnika. Biračima su na raspolaganju dva kandidata A i B. Kako bismo procijenili udio birača koji će glasati za kandidata A, ispitano je n birača. Njih n_1 se izjasnilo za kandidata A, a n_2 za kandidata B ($n = n_1 + n_2$). Neka su slučajne varijable $X_i, i = 1, ..., n$ dane sa

$$X_i = \begin{cases} 1, & i\text{-ti ispitanik se izjasnio za kandidata } A \\ 0, & i\text{-ti ispitanik se izjasnio za kandidata } B. \end{cases}$$

Uzimajući da je ukupna populacija puno veća od veličine uzorka n, možemo pretpostaviti da $X_1, ..., X_n$ slijede Bernoullijevu distribuciju $B(p)$, gdje je p nepoznati parametar koji poprima vrijednosti u $[0, 1]$. Kako bismo proveli bayesovsku analizu, pretpostavljamo da se radi o slučajnoj varijabli. To predstavlja udio birača koji će glasovati za kandidata A u cjelokupnoj populaciji (upravo ono što nas zanima). Dakle,

$$f_{X_i|p}(x \mid p_0) = \begin{cases} p_0, & x = 1 \\ 1 - p_0, & x = 0 \\ 0, & \text{inače} \end{cases}, \quad x \in \mathbb{R},\ p_0 \in [0, 1].$$

Za apriornu razdiobu uzmimo $p \sim Beta(a, b)$, gdje su $a, b > 0$ neki fiksni hiperparametri (v. primjer 1.1.9). Kako bi se naglasila funkcionalna razlika između parametra p, koji je slučajna varijabla i čiju distribuciju procjenjujemo, i pomoćnih fiksnih parametara $a, b \in \mathbb{R}$, u literaturi se susreće izraz **hiperparametri**. Izbor konkretnih vrijednosti za hiperparametre komentirat ćemo na kraju primjera. Uobičajeno je istaknuti da se radi o apriornoj distribuciji oznakom $p_{\text{apriori}} \sim Beta(a, b)$. Dakle,

$$f_p(p_0) = \begin{cases} \frac{1}{B(a,b)} \cdot p_0^{a-1}(1 - p_0)^{b-1}, & 0 \leq p_0 \leq 1 \\ 0, & \text{inače} \end{cases}, \quad p_0 \in \mathbb{R}.$$

gdje je $B(a, b)$ dano sa (1.5). Budući u uzorku $\mathbf{x} = (x_1, ..., x_n)$ točno n_1 podataka poprima vrijednost 1, a njih n_2 vrijednost 0, iz formule (2.5), za $p_0 \in [0,1]$ dobivamo

$$
\begin{aligned}
f_{p|\mathbb{X}}(p_0 \,|\, \mathbf{x}) &\propto f_p(p_0) \cdot \prod_{i=1}^{n} f_{X_i|p}(x_i \,|\, p_0) \\
&\propto p_0^{a-1} \cdot (1-p_0)^{b-1} \cdot \left(p_0^{n_1} \cdot (1-p_0)^{n_2}\right) \\
&\propto p_0^{a+n_1-1} \cdot (1-p_0)^{b+n_2-1}.
\end{aligned}
$$

Uočimo da nije važan poredak pojavljivanja jedinica i nula u uzorku $\mathbf{x}$, već samo njihov broj. Odavde čitamo da je

$$f_{p|\mathbb{X}}(p_0 \,|\, \mathbf{x}) = \frac{1}{B(a+n_1, b+n_2)} \cdot p_0^{a+n_1-1} \cdot (1-p_0)^{b+n_2-1} \ ,$$

jer se upravo za konstantu $1/B(a+n_1, b+n_2)$ dobiva

$$\int_0^1 \frac{1}{B(a+n_1, b+n_2)} \cdot p_0^{a+n_1-1} \cdot (1-p_0)^{b+n_2-1} \ dp_0 = 1.$$

Za $p_0 \notin [0,1]$ možemo staviti $f_{p|\mathbb{X}}(p_0 \,|\, \mathbf{x}) = 0$. Dakle, aposteriorna distribucija od p je $p_{\text{posteriori}} \sim Beta(a+n_1, b+n_2)$. Kako smo u primjeru 1.1.9 pokazali čemu je jednako očekivanje beta distribucije, vidimo da je u ovom slučaju bayesovski procjenitelj jednak

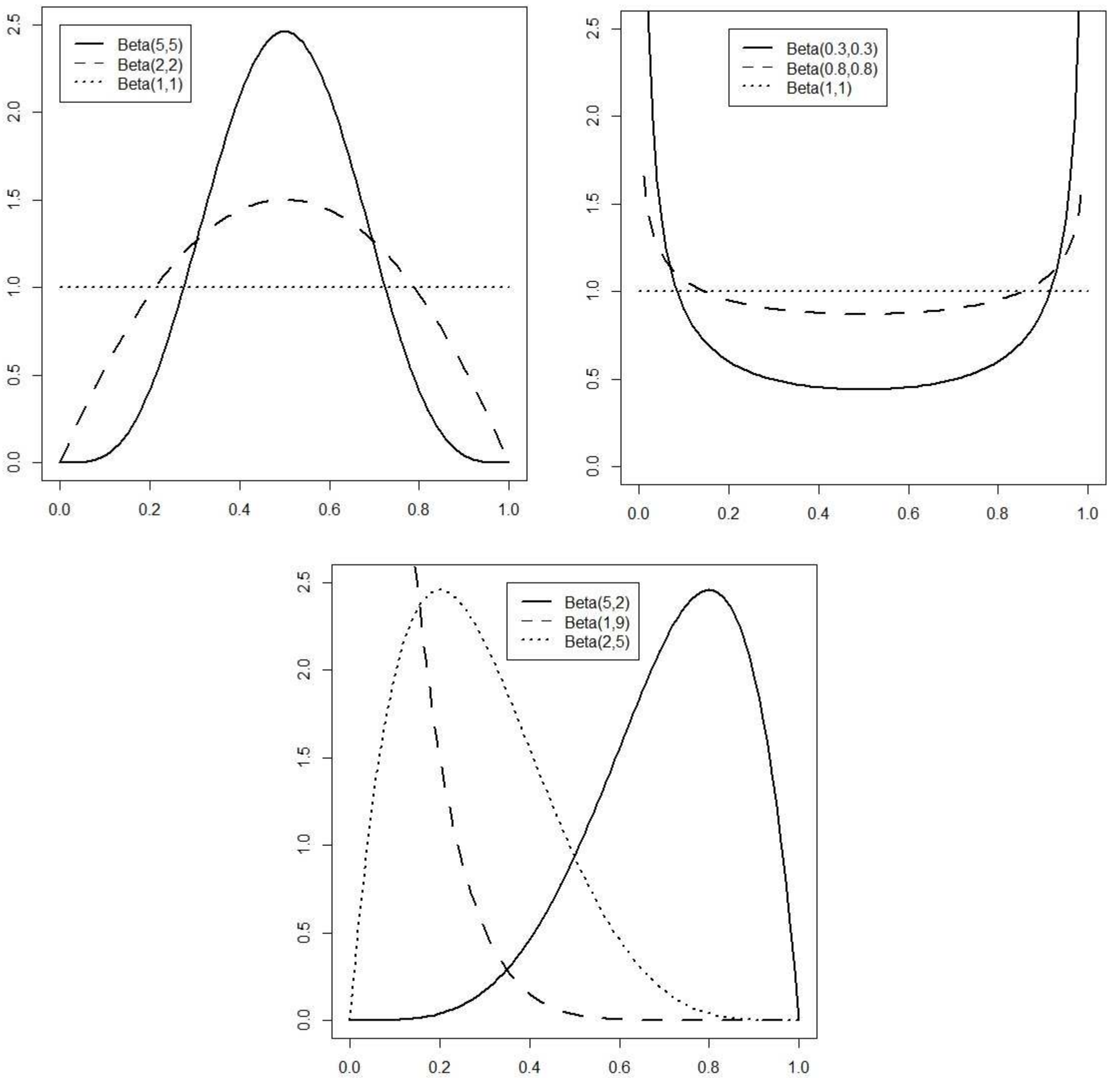
$$\hat{p}_B = \frac{a+n_1}{a+n_1+b+n_2}.$$

Za ovakvu klasu vjerojatnosnih distribucija $\mathcal{P}_0$, koja ima svojstvo (eng. *conjugancy*)

$$p_{\text{apriori}} \in \mathcal{P}_0 \implies p_{\text{posteriori}} \in \mathcal{P}_0 \ ,$$

kažemo da je **konjugirana klasa distribucija** za parametar p slučajnoga uzorka s gustoćama $f_{X_i|p}$ (eng. *sampling distribution*). Dakle, beta distribucije čine konjugiranu klasu distribucija za parametar $B(p)$ distribucije. Konjugirane klase distribucija su u bayesovskoj statistici poželjne jer se na taj način dobivena aposteriorna distribucija može iskoristiti kao apriorna u sljedećem istraživanju.

Preostaje nam komentirati moguće izbore za hiperparametre $a, b > 0$. Na slici 2.1 vidimo na koje sve načine hiperparametri a, b oblikuju funkciju gustoće $Beta(a, b)$ distribucije. U našem se slučaju korisno sjetiti značenja brojeva n_1 i n_2. Budući smo, krečući od $p_{\text{apriori}} \sim Beta(a, b)$, došli do $p_{\text{posteriori}} \sim Beta(a+n_1, b+n_2)$, hiperparametrima a i b pridružujemo značenje broja fiktivnih sudionika u ispitivanju koji su se opredijelili za kandidata A, odnosno kandidata B. Ti brojevi mogu predstavljati naše subjektivno mišljenje,



Slika 2.1: Grafovi funkcija gustoća raznih *beta* distribucija

koje prethodi anketiranju, o tome koliko bi se birača moglo opredijeliti za kojeg kandidata, ali hiperparametri a, b mogu biti odabrani i kao rezultat nekog prethodnog istraživanja. Ukoliko ne raspolažemo takvim podacima i u model ne želimo unositi subjektivna apriorna mišljenja, možemo pokušati sa $p_{\text{apriori}} \sim Beta(1, 1)$. Uočimo da se gustoća $Beta(1, 1)$

distribucije pojednostavljuje na

$$f(p_0) = \begin{cases} 1, & p_0 \in [0, 1] \\ 0, & \text{inače} \end{cases}, \quad p_0 \in \mathbb{R}.$$

Takvu distribuciju nazivamo još i **uniformnom**, odnosno **uniformnom na intervalu** $[0, 1]$, u oznaci $U(0, 1)$. Kao što je prikazano na slici 2.1, ona pridružuje jednaku (uniformnu) gustoću svakoj mogućoj vrijednosti $p_0 \in [0, 1]$ parametra p. U bayesovskoj statistici takve apriorne razdiobe, koje ne favoriziraju niti jednu konkretnu vrijednost parametra, nazivamo **neinformativne apriorne razdiobe** (eng. *noninformative prior*). Za kraj uočimo i da odabirući velike a, b (relativno na n_1, n_2) povećavamo značaj apriornih pretpostavki u odnosu na informacije dobivene anketiranjem.

U idućem su primjeru i nepoznati parametar, kao i elementi slučajnoga uzorka, neprekidne slučajne varijable.

Primjer 2.3.2. Za neprekidnu slučajnu varijablu X kažemo da ima **normalnu razdiobu s parametrima** $\mu \in \mathbb{R}$ i $\sigma^2 > 0$, ako joj je gustoća dana sa

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Pišemo $X \sim N(\mu, \sigma^2)$. Ukoliko je $X \sim N(0, 1)$, govorimo o **jediničnoj normalnoj razdiobi**. Pretpostavimo da imamo (bayesovski) slučajni uzorak $X_1, ..., X_n$, duljine $n \in \mathbb{N}$, iz $N(\mu, \sigma^2)$ razdiobe, gdje je $\sigma^2 > 0$ poznat i fiksan, a μ je nepoznati parametar. Pretpostavljamo da je μ slučajna varijabla te za apriornu distribuciju uzimamo $\mu_{\text{apriori}} \sim N(\theta, \tau^2)$. Hiperparametri $\theta \in \mathbb{R}$ i $\tau^2 > 0$ su također poznati i fiksni. Koristeći Bayesovu formulu (2.5) računamo aposteriornu distribuciju za μ. Neka su $\mu_0, x_1, ..., x_n \in \mathbb{R}$ te $\mathbf{x} = (x_1, ..., x_n)$ i $\bar{x}_n = (x_1 + ... + x_n)/n$. U računu u kojem umjesto jednakosti gledamo proporcionalnosti, možemo zanemariti sve faktore koji ne ovise o μ_0.

$$
\begin{aligned}
f_{\mu|\mathbb{X}}(\mu_0 \,|\, \mathbf{x}) &\propto f_\mu(\mu_0) \cdot f_{\mathbb{X}|\mu}(\mathbf{x} \,|\, \mu_0) \\
&\propto \frac{1}{\cancel{\tau\sqrt{2\pi}}} \cdot e^{-\frac{(\mu_0-\theta)^2}{2\tau^2}} \cdot \prod_{i=1}^{n} \left(\frac{1}{\cancel{\sigma\sqrt{2\pi}}} \cdot e^{-\frac{(x_i-\mu_0)^2}{2\sigma^2}} \right) \\
&\propto \exp\left(-\frac{(\mu_0-\theta)^2}{2\tau^2} - \sum_{i=1}^{n} \frac{(x_i-\mu_0)^2}{2\sigma^2} \right) \\
&\propto \exp\left(-\frac{\sigma^2(\mu_0^2 - 2\mu_0\theta + \theta^2) + \tau^2 \sum_{i=1}^{n}(x_i^2 - 2\mu_0 x_i + \mu_0^2)}{2\tau^2\sigma^2} \right) \\
&\propto \exp\left(-\frac{\mu_0^2\sigma^2 - 2\mu_0\sigma^2\theta - 2\mu_0\tau^2 n\bar{x}_n + \mu_0^2\tau^2 n}{2\tau^2\sigma^2} \right) \cdot \cancel{exp\left(-\frac{\sigma^2\theta^2 + \tau^2\sum_{i=1}^{n} x_i^2}{2\tau^2\sigma^2} \right)} \\
&\propto \exp\left(-\frac{(\sigma^2 + \tau^2 n)\mu_0^2 - 2\mu_0(\sigma^2\theta + \tau^2 n\bar{x}_n)}{2\tau^2\sigma^2} \right) \\
&\propto \exp\left(-\frac{\mu_0^2 - 2\mu_0\frac{\sigma^2\theta+\tau^2 n\bar{x}_n}{\sigma^2+\tau^2 n}}{\frac{2\tau^2\sigma^2}{\sigma^2+\tau^2 n}} \right) \cdot \exp\left(\frac{\left(\frac{\sigma^2\theta+\tau^2 n\bar{x}_n}{\sigma^2+\tau^2 n}\right)^2}{\frac{2\tau^2\sigma^2}{\sigma^2+\tau^2 n}} \right) \propto \exp\left(-\frac{\left(\mu_0 - \frac{\sigma^2\theta+\tau^2 n\bar{x}_n}{\sigma^2+\tau^2 n}\right)^2}{2\cdot\frac{\tau^2\sigma^2}{\sigma^2+\tau^2 n}} \right).
\end{aligned}
$$

Odavde čitamo

$$
\mu_{\text{posteriori}} \sim N\left(\frac{\sigma^2\theta + \tau^2 n\bar{x}_n}{\sigma^2 + \tau^2 n}, \ \left(\frac{\tau\sigma}{\sqrt{\sigma^2 + \tau^2 n}} \right)^2 \right).
$$

Dakle, pokazali smo da normalne distribucije čine konjugiranu klasu distribucija za očekivanje μ uzorka iz normalne distribucije $N(\mu, \sigma^2)$ s poznatom varijancom σ^2.

Ovime smo pokrili osnove parametarske bayesovske statistike. Idući korak bilo bi izučavanje višeparametarskih modela te asimptotskih rezultata u parametarskoj bayesovskoj statistici. Time se u ovome radu nećemo baviti, već zainteresiranoga čitatelja upućujemo na [2], poglavlja 3 i 4.

Poglavlje 3

Dirichletov proces

U ovom poglavlju uvodimo središnji pojam ovoga rada, a to je *Dirichletov proces*. Teorija će biti izložena po uzoru na [1] i [2]. Primjenu Dirichletovog procesa u tzv. *neparametarskoj bayesovskoj statistici* ilustrirat ćemo kroz nekoliko primjera i pokazat ćemo kako simulirati podatke iz Dirichletovog procesa.

3.1 Motivacija

Slično kao u poglavlju 2, promatramo realizaciju $x_1, ..., x_n \in \mathbb{R}$ $(n \in \mathbb{N})$ slučajnoga uzorka $X_1, ..., X_n$ iz nepoznate distribucije. Tada smo pretpostavljali da uzorak dolazi iz neke *parametrizirane familije distribucija*, npr. $X_1, ..., X_n \sim B(p)$ ili $N(\mu, \sigma^2)$, pri čemu je p, odnosno μ nepoznati parametar. Pretpostavljajući dodatno da je nepoznati parametar slučajna varijabla, daljnja se analiza svodila na određivanje njegove aposteriorne distribucije. Pri tome je $X_1, ..., X_n$ bio slučajni uzorak u bayesovskom smislu, odnosno slučajne varijable $X_1, ..., X_n$ su bile nezavisne i jednako distribuirane uvjetno na promatrani parametar. Na temelju uzorka $x_1, ..., x_n$ sada želimo predložiti distribuciju koju bi mogle slijediti $X_1, ..., X_n$, pri čemu u obzir dolaze sve moguće vjerojatnosne distribucije. Drugim riječima, ne ograničavamo se na neku parametriziranu familiju. Uočimo da je $X_1, ..., X_n$ sada slučajni uzorak u standardnom smislu, odnosno radi se o nezavisnim i jednako distribuiranim slučajnim varijablama.

Neka je, dakle, dana slučajna varijabla X na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. U odjeljku 1.3 komentirali smo kako odrediti distribuciju od X zapravo znači odrediti vjerojatnosnu mjeru $\mathbb{P}_X : \mathcal{B} \to \mathbb{R}$, danu sa

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \quad B \in \mathcal{B}.$$

Na temelju realizacija $x_1, ..., x_n \in \mathbb{R}$ nezavisnih slučajnih varijabli $X_1, ..., X_n$, koje su sve distribuirane jednako kao X, želimo zaključiti nešto o $\mathbb{P}_X$. Za razliku od klasične (frekvencionističke) statistike, a slično kao i u poglavlju 2, ne želimo dati procjenu za $\mathbb{P}_X$ u obliku samo jedne konkretne vjerojatnosne mjere koja najbolje odgovara podacima $x_1, ..., x_n$. Ako sa $\mathcal{P}$ označimo familiju svih mogućih vjerojatnosnih mjera na izmjerivom prostoru $(\mathbb{R}, \mathcal{B})$, kao rezultat analize želimo dobiti neku vjerojatnosnu distribuciju nad skupom (vjerojatnosnih distribucija) $\mathcal{P}$. Pokazuje se da su takvi objekti istovjetni slučajnim procesima indeksiranima po skupu $T = \mathcal{B}$. Zaista, pretpostavimo da je $\mathbb{P}_X$ slučajni proces na vjerojatnosnom prostoru $(\mathbb{R}^T, \mathcal{B}_T, \mathbb{P}_T)$ sa skupom indeksa $T = \mathcal{B}$, kao što je to opisano u odjeljku 1.3. Tada je za svako $\omega \in \mathbb{R}^T$ realizacija $\mathbb{P}_X(\omega) \in \mathbb{R}^T$, odnosno funkcija $\mathcal{B} \to \mathbb{R}$. Važno je uočiti da nisu sve funkcije $\mathcal{B} \to \mathbb{R}$ vjerojatnosne mjere na $(\mathbb{R}, \mathcal{B})$. Drugim riječima, inkluzija $\mathcal{P} \subset \mathbb{R}^{\mathcal{B}}$ je stroga. Međutim, mi ćemo zahtijevati da je

$$\mathbb{P}_X \in \mathcal{P} \quad (g.s.). \tag{3.1}$$

Takav slučajni proces $\mathbb{P}_X$ tada nazivamo **slučajnom vjerojatnosnom mjerom**, odnosno **slučajnom vjerojatnošću**. U literaturi se često spominje i pojam **slučajne mjere** (eng. *random measure*) pri čemu je uvjet (3.1) zamijenjen sa

$$\mathbb{P}_X \in \mathcal{G} \quad (g.s.),$$

gdje $\mathcal{G}$ označava familiju svih mjera na izmjerivom prostoru $(\mathbb{R}, \mathcal{B})$.

Primjer 3.1.1. Neka je $T = \mathcal{B}$. Promatrajmo izmjerivi prostor $(\mathbb{R}^T, \mathcal{B}_T)$, opisan u odjeljku 1.3. Neka je $\mathbb{P}_0 \in \mathcal{P}$ neka konkretna vjerojatnosna mjera na $(\mathbb{R}, \mathcal{B})$. Neka je $\mathbb{P}_T$ vjerojatnost na $(\mathbb{R}^T, \mathcal{B}_T)$ takva da za svako $A \in \mathcal{B}_T$ vrijedi

$$\mathbb{P}_T(A) = \delta_{\mathbb{P}_0}(A) = \begin{cases} 1, & \mathbb{P}_0 \in A \\ 0, & \mathbb{P}_0 \notin A. \end{cases}$$

Sjetimo se, elementi od $\mathcal{B}_T$ su podskupovi skupa $\mathbb{R}^T \supset \mathcal{P}$ pa se za $A \in \mathcal{B}_T$ ima smisla pitati je li $\mathbb{P}_0 \in A$. Tako definirana $\mathbb{P}_T$ je očito vjerojatnost na $(\mathbb{R}^T, \mathcal{B}_T)$, a funkcija $\mathbb{P}_X : \mathbb{R}^T \to \mathbb{R}^T$, dana sa $\mathbb{P}_X(\omega) = \omega$, $\forall \omega \in \mathbb{R}^T$, je onda slučajni proces. Štoviše, kako je $\mathbb{P}_0 \in \mathcal{P}$, vrijedi i

$$\mathbb{P}_T\left(\mathbb{P}_X \in \mathcal{P}\right) = \mathbb{P}_T\left(\left\{\omega \in \mathbb{R}^T : \mathbb{P}_X(\omega) \in \mathcal{P}\right\}\right) =$$

$$= \mathbb{P}_T\left(\left\{\omega \in \mathbb{R}^T : \omega \in \mathcal{P}\right\}\right) = \mathbb{P}_T\left(\mathcal{P}\right) = 1.$$

Dakle, $\mathbb{P}_X \in \mathcal{P}$ $(g.s.)$, odnosno $\mathbb{P}_X$ je slučajna vjerojatnost. Možemo reći da je ovako zadana $\mathbb{P}_X$ degenerirana slučajna vjerojatnost, budući ona gotovo sigurno poprima vrijednost upravo $\mathbb{P}_0$ (slično kao što i slučajne varijable koje gotovo sigurno poprimaju neku konstantnu realnu vrijednost zovemo degeneriranima).

Vratimo se na polazni problem. Ovdje također želimo primijeniti osnovnu ideju bayesovskog zaključivanja. Krećemo od neke apriorne slučajne vjerojatnosti $\mathbb{P}_X$. Zatim opažamo podatke $(x_1, ..., x_n) \in \mathbb{R}^n$, koji su realizacija slučajnoga uzorka $\mathbb{X} = (X_1, ..., X_n)$. Koristeći Bayesovu shemu

$$\textit{posterior} \propto \textit{prior} \cdot \textit{likelihood},$$

dolazimo do aposteriorne slučajne vjerojatnosti $\mathbb{P}_X \,|\, \mathbb{X}\,(\,\cdot\,|\, x_1, ..., x_n)$. Pokazuje se da je za apriornu slučajnu vjerojatnost $\mathbb{P}_X$ pogodno uzeti tzv. *Dirichletov proces*. Njegova će definicija biti bazirana na Kolmogorovljevim uvjetima suglasnosti i teoremu 1.3.5. Konačnodimenzionalne razdiobe tog procesa zovemo *Dirichletovim razdiobama*, tj. distribucijama.

3.2 Dirichletova distribucija

Promatramo slučajnu vjerojatnost $\mathbb{P}_X$. Neka su uz nju dani $k \in \mathbb{N}$, $k \geq 2$ te $\{B_1, ..., B_k\} \subset \mathcal{B}$ fiksna izmjeriva particija skupa $\mathbb{R}$. Uočimo da je tada $(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_k))$ k-dimenzionalni slučajni vektor koji poprima vrijednosti u skupu

$$H_k^* = \left\{\pi = (\pi_1, ...\pi_k) \in \mathbb{R}^k : \pi_1, ..., \pi_k \geq 0, \sum_{j=1}^{k} \pi_j = 1\right\} \subset \mathbb{R}^k. \tag{3.2}$$

Budući je

$$\mathbb{P}_X(B_k) = 1 - (\mathbb{P}_X(B_1) + ... + \mathbb{P}_X(B_{k-1})), \tag{3.3}$$

promatrat ćemo $(k-1)$-dimenzionalni slučajni vektor $(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_{k-1}))$, koji pak poprima vrijednosti u skupu

$$H_{k-1} = \left\{(\pi_1, ..., \pi_{k-1}) \in \mathbb{R}^{k-1} : \pi_1, ..., \pi_{k-1} \geq 0, \sum_{j=1}^{k-1} \pi_j \leq 1\right\} \subset \mathbb{R}^{k-1}. \tag{3.4}$$

Definicija 3.2.1. *Neka je $k \in \mathbb{N}$, $k \geq 2$. Za $(k-1)$-dimenzionalni neprekidni slučajni vektor* ***P*** *kažemo da ima* ***Dirichletovu distribuciju*** *s parametrima $a_1, ..., a_k > 0$, ako mu je gustoća dana sa*

$$f(\pi_1, ..., \pi_{k-1}\,;\, a_1, ..., a_k) = \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right)}{\prod_{j=1}^{k} \Gamma(a_j)} \cdot \prod_{j=1}^{k} \pi_j^{a_j - 1}, \tag{3.5}$$

za svako $(\pi_1, ..., \pi_{k-1}) \in H_{k-1}$, gdje je H_{k-1} dan sa (3.4)*, dok je $f(\pi_1, ..., \pi_{k-1}\,;\, a_1, ..., a_k) = 0$, za $(\pi_1, ..., \pi_{k-1}) \notin H_{k-1}$. Funkcija Γ dana je sa* (1.6)*. Pri tome podrazumijevamo*

$$\pi_k = 1 - (\pi_1 + ... + \pi_{k-1}). \tag{3.6}$$

Stoga funkciju $f(\,\cdot\,; a_1, ..., a_k)$ iz (3.5) *shvaćamo kao funkciju argumenta $\pi = (\pi_1, ..., \pi_k)$. Pišemo* ***P*** *$\sim$ Dirichlet$(a_1, ..., a_k)$, ali koristi se i oznaka $f \sim$ Dirichlet$(a_1, ..., a_k)$. Poseban slučaj takve distribucije je za $a_1 = ... = a_k = a > 0$, kada se formula* (3.5) *svodi na*

$$f(\pi\,;\, a) \equiv f(\pi\,;\, a, ..., a) = \frac{\Gamma(ka)}{(\Gamma(a))^k} \cdot \prod_{j=1}^{k} \pi_j^{a-1}. \tag{3.7}$$

Tada pišemo ***P*** *$\sim$ Dirichlet(a), odnosno $f \sim$ Dirichlet(a).*

Uočimo da je Dirichletova distribucija generalizacija beta distribucije, koja se dobiva u slučaju $k = 2$ (v. primjer 1.1.9).

Funkcija $f \sim$ *Dirichlet*$(a_1, ..., a_k)$ je očito izmjeriva i nenegativna, ali preostaje vidjeti da vrijedi

$$\int_{H_{k-1}} f(\pi_1, ..., \pi_k\,;\, a_1, ..., a_k)\, d\lambda^{k-1}(\pi_1, ..., \pi_{k-1}) = 1. \tag{3.8}$$

Lema 3.2.2. *Za svako $k \in \mathbb{N}$, $k \geq 2$ i proizvoljne $a_1, ..., a_k > 0$ vrijedi* (3.8)*, gdje je $f \sim$ Dirichlet$(a_1, ..., a_k)$, a H_{k-1} je dan sa* (3.4).

Dokaz. Jednakost pokazujemo matematičkom indukcijom po $k \geq 2$. Za $k = 2$ i $a_1, a_2 > 0$ proizvoljne imamo

$$f(\pi_1, \pi_2\,;\, a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1) \cdot \Gamma(a_2)} \cdot \pi_1^{a_1 - 1} \cdot (1 - \pi_1)^{a_2 - 1} = \left\{(1.7)\right\}$$

$$= \frac{1}{B(a_1, a_2)} \cdot \pi_1^{a_1 - 1} \cdot (1 - \pi_1)^{a_2 - 1}$$

što je upravo gustoća $Beta(a_1, a_2)$ distribucije (v. primjer 1.1.9) pa u tom slučaju već znamo da vrijedi

$$\int_{H_1} f(\pi_1, \pi_2\,;\, a_1, a_2)\, d\lambda(\pi_1) = \int_0^1 f(\pi_1, \pi_2\,;\, a_1, a_2)\, d\pi_1 = 1.$$

Pretpostavimo sada da jednakost (3.8) vrijedi za neko $k \geq 2$ i proizvoljne $a_1, ..., a_k > 0$. Za $f \sim Dirichlet(a_1, ..., a_{k+1})$, računamo

$$\int_{H_k} f(\pi_1, ...\pi_k, \pi_{k+1}\,;\, a_1, ..., a_{k+1})\ \ d\lambda^k(\pi_1, ..., \pi_k) =$$

$$= \int_{H_k} \frac{\Gamma\left(\sum_{j=1}^{k+1} a_j\right)}{\prod_{j=1}^{k+1} \Gamma(a_j)} \cdot \left(\prod_{j=1}^{k} \pi_j^{a_j - 1}\right) \cdot (1 - (\pi_1 + ... + \pi_k))^{a_{k+1}-1}\ \ d\lambda^k(\pi_1, ..., \pi_k) = I_1.$$

Provest ćemo sljedeću zamjenu varijabli:

$$\begin{cases} \tilde{\pi}_1 & = \frac{\pi_1}{\pi_1 + ... + \pi_k} \\ & \vdots \\ \tilde{\pi}_{k-1} & = \frac{\pi_{k-1}}{\pi_1 + ... + \pi_k} \\ \tilde{\pi}_k & = \pi_1 + ... + \pi_k \end{cases}$$

odnosno

$$g(\pi_1, ...\pi_k) = \left(\frac{\pi_1}{\pi_1 + ... + \pi_k}, ..., \frac{\pi_{k-1}}{\pi_1 + ... + \pi_k}, \pi_1 + ... + \pi_k\right).$$

Lako se vidi da je

$$g^{-1}(\tilde{\pi}_1, ...\tilde{\pi}_k) = (\tilde{\pi}_1\tilde{\pi}_k, ..., \tilde{\pi}_{k-1}\tilde{\pi}_k, \tilde{\pi}_k - \tilde{\pi}_k(\tilde{\pi}_1 + ... + \tilde{\pi}_{k-1}))$$

pa je Jacobijan $Dg^{-1}(\tilde{\pi}_1, ...\tilde{\pi}_k) =$

$$= \begin{vmatrix} \tilde{\pi}_k & 0 & ... & 0 & \tilde{\pi}_1 \\ 0 & \tilde{\pi}_k & ... & 0 & \tilde{\pi}_2 \\ & & \vdots & & \\ 0 & 0 & ... & \tilde{\pi}_k & \tilde{\pi}_{k-1} \\ -\tilde{\pi}_k & -\tilde{\pi}_k & ... & -\tilde{\pi}_k & 1 - (\tilde{\pi}_1 + ... + \tilde{\pi}_{k-1}) \end{vmatrix} = \begin{vmatrix} \tilde{\pi}_k & 0 & ... & 0 & \tilde{\pi}_1 \\ 0 & \tilde{\pi}_k & ... & 0 & \tilde{\pi}_2 \\ & & \vdots & & \\ 0 & 0 & ... & \tilde{\pi}_k & \tilde{\pi}_{k-1} \\ 0 & 0 & ... & 0 & 1 \end{vmatrix} = \tilde{\pi}_k^{k-1}.$$

Naime, zadnjem smo retku polazne matrice dodali gornjih $k-1$ redaka, nakon čega smo dobili gornje-trokutastu matricu s nepromijenjenom determinantom. Novi skup po kojem integriramo je

$$\left\{(\tilde{\pi}_1, ...\tilde{\pi}_k) \in \mathbb{R}^k : \tilde{\pi}_1, ..., \tilde{\pi}_{k-1} \geq 0, \sum_{j=1}^{k-1} \tilde{\pi}_j \leq 1, \tilde{\pi}_k \in [0, 1]\right\} = H_{k-1} \times [0, 1].$$

Stoga dobivamo

$$I_1 = \int_{H_{k-1}} \int_0^1 \frac{\Gamma\left(\sum_{j=1}^{k+1} a_j\right)}{\prod_{j=1}^{k+1} \Gamma(a_j)} \cdot \left(\prod_{j=1}^{k-1} (\tilde{\pi}_j \tilde{\pi}_k)^{a_j-1}\right) \cdot \left(\tilde{\pi}_k - \tilde{\pi}_k \cdot \sum_{j=1}^{k-1} \tilde{\pi}_j\right)^{a_k-1} \cdot (1-\tilde{\pi}_k)^{a_{k+1}-1} \cdot \left|\tilde{\pi}_k^{k-1}\right| \\ d\tilde{\pi}_k \, d\lambda^{k-1}(\tilde{\pi}_1, ..., \tilde{\pi}_{k-1}) =$$

$$= \left\{ \text{iskoristimo relaciju(1.7) i uočimo da je } \tilde{\pi}_k \geq 0 \right\} =$$

$$= \int_{H_{k-1}} \int_0^1 \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right) \Gamma(a_{k+1})}{B(a_1 + ... + a_k, a_{k+1}) \cdot \prod_{j=1}^{k+1} \Gamma(a_j)} \cdot \left(\prod_{j=1}^{k-1} \tilde{\pi}_j^{a_j-1}\right) \cdot \tilde{\pi}_k^{a_1+...+a_{k-1}-k+1} \cdot \\ \cdot \tilde{\pi}_k^{a_k-1} \cdot \left(1 - \sum_{j=1}^{k-1} \tilde{\pi}_j\right)^{a_k-1} \cdot (1-\tilde{\pi}_k)^{a_{k+1}-1} \cdot \tilde{\pi}_k^{k-1} \;\; d\tilde{\pi}_k \;\; d\lambda^{k-1}(\tilde{\pi}_1, ..., \tilde{\pi}_{k-1}) =$$

$$= \left(\int_0^1 \frac{1}{B(a_1 + ... + a_k, a_{k+1})} \cdot \tilde{\pi}_k^{a_1+...+a_k-1} (1-\tilde{\pi}_k)^{a_{k+1}-1} \;\; d\tilde{\pi}_k \right) \cdot \\ \cdot \left(\int_{H_{k-1}} \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right)}{\cdot \prod_{j=1}^{k} \Gamma(a_j)} \cdot \left(\prod_{j=1}^{k-1} \tilde{\pi}_j^{a_j-1}\right) \cdot \left(1 - \sum_{j=1}^{k-1} \tilde{\pi}_j\right)^{a_k-1} \;\; d\lambda^{k-1}(\tilde{\pi}_1, ..., \tilde{\pi}_{k-1}) \right) = 1 \cdot 1 = 1,$$

gdje je prvi integral u gornjem izrazu jednak jedinici jer se radi o integralu gustoće $Beta(a_1 + ... + a_k, a_{k+1})$ razdiobe, a drugi po pretpostavci indukcije. Time smo pokazali (3.8) za svaki izbor $k \geq 2$ te $a_1, ..., a_k > 0$.

□

Dakle, funkcija $\pi \mapsto f(\pi\,; a_1, ..., a_k)$, dana sa (3.5), je funkcija gustoće nekog $(k-1)$-dimenzionalnog neprekidnog slučajnog vektora, koji poprima vrijednosti u skupu H_{k-1}, danog sa (3.4).

Napomena 3.2.3. Uočimo da postoji konkretan razlog zašto smo sa $(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_k))$ morali prijeći na vektor $(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_{k-1}))$. Naime, prvi vektor poprima vrijednosti u skupu H_k^*, danim sa (3.2). Radi se o podskupu $(k-1)$-dimenzionalne plohe (točnije o $(k-1)$-dimenzionalnom simpleksu) u k-dimenzionalnom skupu $\mathbb{R}^k$. Stoga bi za integral bilo koje izmjerive funkcije $f : \mathbb{R}^k \to \mathbb{R}$ nužno vrijedilo

$$\int_{H_k^*} f \;\; d\lambda^k = 0.$$

Dakle, ne postoji način da za $(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_k))$ definiramo gustoću neprekidnog slučajnog vektora. S druge strane, vidjeli smo da to možemo učiniti za $(k-1)$-dimenzionalni slučajni vektor $(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_{k-1}))$. Uzimajući u obzir (3.3), mi ćemo poistovjećivati ta dva slučajna vektora. Formalno, uzimat ćemo da se radi o $(k-1)$-dimenzionalnom slučajnom vektoru koji poprima vrijednosti u H_{k-1}, danim sa (3.4).

Primjer 3.2.4. Neka je $\mathbb{P}_X$ slučajna vjerojatnost i $\{B_1, ..., B_k\} \subset \mathcal{B}$ izmjeriva particija skupa $\mathbb{R}$, pri čemu je $k \geq 2$. Pretpostavimo $(\mathbb{P}_X(B_1), ..., \mathbb{P}_{\mathbb{X}}(B_k)) \sim \mathit{Dirichlet}(a_1, ..., a_k)$, za neke $a_1, ..., a_k > 0$. To smatramo apriornom distribucijom tog slučajnog vektora. Koristimo oznaku $\mathbf{P} = (\mathbb{P}_X(B_1), ..., \mathbb{P}_{\mathbb{X}}(B_k))$. Fiksirajmo $(\pi_1, ..., \pi_{k-1}) \in H_{k-1}$ te $\pi_k = 1 - (\pi_1 + ... + \pi_{k-1})$. Neka je $\mathbb{X} = (X_1, ..., X_n)$ slučajni uzorak za X te $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ jedna njegova realizacija. Označimo za $i = 1, ..., n$

$$Y_i = j, \quad \text{ako je } X_i \in B_j, \quad j = 1, ..., k.$$

Slučajna varijabla Y_i bilježi u koji skup particije $\{B_1, ..., B_k\}$ je upala X_i. Uočimo da Y_i, uvjetno na poznavanje slučajnoga vektora $\mathbf{P}$, ima sljedeću diskretnu distribuciju

$$Y_i \,|\, \mathbf{P} \sim \begin{pmatrix} 1 & ... & k \\ \pi_1 & ... & \pi_k \end{pmatrix}.$$

Štoviše, $Y_1, ..., Y_n$ su nezavisne i jednako distribuirane uvjetno na $\mathbf{P}$. Označimo sa $\mathbb{Y} = (Y_1, ..., Y_n)$ te neka je $\mathbf{y} = (y_1, ..., y_n)$ pripadna realizacija od $\mathbb{Y}$ (u skladu sa danim $\mathbf{x}$). Razumno je pretpostaviti $f_{\mathbb{Y}}(\mathbf{y}) > 0$. Pretpostavimo da je u uzorku $\mathbf{x}$ točno n_j podataka upalo u skup B_j, odnosno da je među $y_1, ..., y_n$ točno n_j podataka jednako broju j $(j = 1, ..., k)$, pri čemu je $n_1 + ... + n_k = n$. Slično kao u poglavlju 2, možemo iskoristiti formulu (2.5).

$$\begin{aligned}
f_{\mathbf{P}|\mathbb{X}}(\pi \,|\, \mathbf{x}) \equiv f_{\mathbf{P}|\mathbb{Y}}(\pi \,|\, \mathbf{y}) &\propto f_{\mathbf{P}}(\pi) \cdot f_{\mathbb{Y}|\mathbf{P}}(\mathbf{y} \,|\, \pi) \\
&\propto \cancel{\frac{\Gamma\left(\sum_{j=1}^k a_j\right)}{\prod_{j=1}^k \Gamma(a_j)}} \cdot \prod_{j=1}^k \pi_j^{a_j - 1} \cdot \prod_{j=1}^k \pi_j^{n_j} \\
&\propto \prod_{j=1}^k \pi_j^{a_j + n_j - 1} \\
&\propto \frac{\Gamma\left(\sum_{j=1}^k (a_j + n_j)\right)}{\prod_{j=1}^k \Gamma(a_j + n_j)} \cdot \prod_{j=1}^k \pi_j^{a_j + n_j - 1}.
\end{aligned}$$

Zaključujemo da je aposteriorna distribucija od $\mathbf{P}$ ponovno Dirichletova i to $\mathbf{P} \sim Dirichlet(a_1 + n_1, ..., a_k + n_k)$. Dakle, Dirichletove distribucije ovdje čine konjugiranu klasu. Prilikom izbora konkretnih vrijednosti za parametre $a_1, ..., a_k > 0$, možemo se voditi analognim razmišljanjima kao u primjeru 2.3.1, tj. da a_j predstavlja broj fiktivnih podataka koji su sadržani u B_j, $j = 1, ..., k$.

Podsjetimo se još jedne vjerojatnosne distribucije. Za neprekidnu slučajnu varijablu X kažemo da ima **gama razdiobu s parametrima** $\alpha, \beta > 0$, ako joj je gustoća dana sa

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

gdje je funkcija Γ dana sa (1.6). Tako zadana f_X je očito nenegativna, a $\int_0^\infty f_X(x)\,\mathrm{d}x = 1$ lagano slijedi iz (1.6). Pišemo $X \sim Gamma(\alpha, \beta)$.

Neka su $Y_1, ..., Y_k$ ($k \geq 2$) nezavisne slučajne varijable, $Y_j \sim Gamma(a_j, 1)$, $j = 1, ..., k$, za neke $a_1, ..., a_k > 0$. Definiramo

$$P_j = \frac{Y_j}{\sum_{i=1}^k Y_i}, \quad j = 1, ..., k.$$

U [3], Lema 2.1.2., pokazano je da slučajni vektor $(P_1, ..., P_k)$ ima $Dirichlet(a_1, ..., a_k)$ distribuciju te da je nezavisan od slučajne varijable

$$Y = \sum_{i=1}^k Y_i.$$

Lema 3.2.5. *Neka su $Y_1, ..., Y_k$ nezavisne slučajne varijable, $Y_j \sim Gamma(a_j, \beta)$, $j = 1, ..., k$, za neke $a_1, ..., a_k, \beta > 0$. Tada je*

$$Y = \sum_{j=1}^k Y_j \sim Gamma(a_1 + ... + a_k, \beta).$$

Dokaz. Karakteristična funkcija ϕ_j slučajne varijable $Y_j \sim Gamma(a_j, \beta)$ jednaka je

$$\phi_j(t) = (1 - i\beta t)^{-a_j}, \quad t \in \mathbb{R},$$

za svako $j = 1, ..., k$ (v. [5], 13., posebno Primjer 13.1.(k)). Koristeći nezavisnost slučajnih varijabli $Y_1, ..., Y_k$ dobivamo (v. [5], Teorem 13.1.(b)) da je karakteristična funkcija ϕ slučajne varijable Y jednaka

$$\phi(t) = \prod_{j=1}^k \phi_j(t) = (1 - i\beta t)^{-(a_1 + ... + a_k)}, \quad t \in \mathbb{R}.$$

Sada po teoremu jedinstvenosti za karakteristične funkcije (v. [5], Teorem 13.2.) zaključujemo $Y \sim Gamma(a_1 + ... + a_k, \beta)$.

□

Propozicija 3.2.6. *Neka je* $(P_1, ..., P_k) \sim Dirichlet(a_1, ..., a_k)$ $(k \geq 2)$. *Tada je*

$$\left(\sum_{j=1}^{k_1} P_j, \sum_{j=k_1+1}^{k_2} P_j, ..., \sum_{j=k_{l-1}+1}^{k_l} P_j\right) \sim Dirichlet\left(\sum_{j=1}^{k_1} a_j, \sum_{j=k_1+1}^{k_2} a_j, ..., \sum_{j=k_{l-1}+1}^{k_l} a_j\right)$$

za svaki izbor $1 \leq k_1 < k_2 < ... < k_l = k$.

Dokaz. Zbog $(P_1, ..., P_k) \sim Dirichlet(a_1, ..., a_k)$ zaključujemo da postoje nezavisne slučajne varijable $Y_1, ..., Y_k$ takve da je $Y_j \sim Gamma(a_j, 1)$, $j = 1, ..., k$ i

$$P_j = \frac{Y_j}{\sum_{i=1}^{k} Y_i}, \quad j = 1, ..., k.$$

Iz leme 3.2.5 dobivamo

$$\begin{aligned}
\sum_{j=1}^{k_1} Y_j &\sim Gamma\left(\sum_{j=1}^{k_1} a_j, 1\right), \\
\sum_{j=k_1+1}^{k_2} Y_j &\sim Gamma\left(\sum_{j=k_1+1}^{k_2} a_j, 1\right) \\
&\vdots \\
\sum_{j=k_{l-1}+1}^{k_l} Y_j &\sim Gamma\left(\sum_{j=k_{l-1}+1}^{k_l} a_j, 1\right).
\end{aligned}$$

Konačno, označimo li

$$Y = \sum_{i=1}^{k} Y_i = \sum_{i=1}^{k_1} Y_i + \sum_{i=k_1+1}^{k_2} Y_i + ... + \sum_{i=k_{l-1}+1}^{k_l} Y_i,$$

dobivamo

$$\left(\sum_{j=1}^{k_1} P_j, \sum_{j=k_1+1}^{k_2} P_j, ..., \sum_{j=k_{l-1}+1}^{k_l} P_j\right) = \left(\sum_{j=1}^{k_1} \frac{Y_j}{Y}, \sum_{j=k_1+1}^{k_2} \frac{Y_j}{Y}, ..., \sum_{j=k_{l-1}+1}^{k_l} \frac{Y_j}{Y}\right)$$

$$= \left(\frac{\sum_{j=1}^{k_1} Y_j}{Y}, \frac{\sum_{j=k_1+1}^{k_2} Y_j}{Y}, ..., \frac{\sum_{j=k_{l-1}+1}^{k_l} Y_j}{Y}\right)$$

$$\sim \mathit{Dirichlet}\left(\sum_{j=1}^{k_1} a_j, \sum_{j=k_1+1}^{k_2} a_j, ..., \sum_{j=k_{l-1}+1}^{k_l} a_j\right).$$

□

3.3 Dirichletov proces

Sada smo spremni izreći definiciju Dirichletovog procesa. U nastavku ćemo za skupove $B_1, ..., B_m \in \mathcal{B}$ reći da čine izmjerivu particiju skupa $\mathbb{R}$, ukoliko su ti skupovi u parovima disjunktni te u uniji daju čitavi $\mathbb{R}$. Dakle, izostavljamo uvjet $B_j \neq \emptyset$, za svako $j = 1, ..., m$.

Definicija 3.3.1. *Neka je $\alpha > 0$ te $\mathbb{P}_0 \in \mathcal{P}$ neka (konkretna, neslučajna) vjerojatnosna mjera na $(\mathbb{R}, \mathcal{B})$. Za slučajnu vjerojatnost $\mathbb{P}_X$ kažemo da je* ***Dirichletov proces*** *ako za svako $k \in \mathbb{N}$, $k \geq 2$ i za svaku $\{B_1, ..., B_k\} \subset \mathcal{B}$ izmjerivu particiju skupa $\mathbb{R}$, takvu da je $\mathbb{P}_0(B_1), ..., \mathbb{P}_0(B_k) > 0$, vrijedi*

$$(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_k)) \sim \mathit{Dirichlet}\,(\alpha\mathbb{P}_0(B_1), ..., \alpha\mathbb{P}_0(B_k))$$

te ako je $\mathbb{P}_X(B) = 0$ (g.s.), za svako $B \in \mathcal{B}$ takvo da je $\mathbb{P}_0(B) = 0$. U tom slučaju pišemo $\mathbb{P}_X \sim DP\,(\alpha\mathbb{P}_0)$. Parametar α zovemo skalirajućim parametrom (eng. scaling parameter), a $\mathbb{P}_0$ baznom vjerojatnošću (eng. baseline probability measure).

Na prvu nije jasno jesmo li definicijom 3.3.1 uopće zadali sve konačno-dimenzionalne distribucije slučajnoga procesa $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$ niti zadovoljavaju li one Kolmogorovljeve uvjete suglasnosti iz definicije 1.3.4. Kako bismo opravdali egzistenciju tog slučajnog procesa (pozivajući se na teorem 1.3.5), u nastavku otklanjamo ta dva problema. To činimo po uzoru na [1].

Prvenstveno, neka su $k \in \mathbb{N}$ te $A_1, ..., A_k \in \mathcal{B}$ proizvoljni i različiti. Za svaki podskup indeksa $I \subseteq \{1, ..., k\}$, promatramo

$$B_I = \left(\bigcap_{i \in I} A_i\right) \cap \left(\bigcap_{i \notin I} (\mathbb{R} \setminus A_i)\right).$$

U ovom kontekstu uzimamo $\bigcap_{i\in\emptyset} A_i = \mathbb{R}$. Budući imamo 2^k mogućih izbora za I, skupove B_I smijemo indeksirati brojevima $1, ..., 2^k$. Kako su $A_1, ..., A_k$ međusobno različiti, to su $B_1, ..., B_{2^k}$ u parovima disjunktni te u uniji daju čitavi $\mathbb{R}$. U nastavku želimo promatrati samo one B_j za koje je $\mathbb{P}_0(B_j) > 0$. Uočimo da to mora biti zadovoljeno za barem jedno B_j. (U suprotnom bismo imali $\mathbb{P}_0(\mathbb{R}) = \mathbb{P}_0(B_1) + ... + \mathbb{P}_0(B_{2^k}) = 0$). Nije smanjenje općenitosti pretpostaviti da to vrijedi točno za prvih m skupova, tj. $B_1, ..., B_m$ ($m \leq 2^k$). Problem je što ti skupovi sada više ne čine nužno particiju skupa $\mathbb{R}$. (Zaista, može biti da je neko $B_j \neq \emptyset$, ali $\mathbb{P}_0(B_j) = 0$). Ipak, po pretpostavci znamo da je

$$\mathbb{P}_X(B_j) = 0 \quad (g.s.),$$

za $j > m$. Umjesto skupa B_m možemo promatrati

$$\tilde{B}_m = B_m \cup \bigcup_{j=m+1}^{2^k} B_j.$$

Pri tome je očito

$$\mathbb{P}_0(\tilde{B}_m) = \mathbb{P}_0(B_m) + \mathbb{P}_0(B_{m+1}) + ... + \mathbb{P}_0(B_{2^k}) = \mathbb{P}_0(B_m),$$

a kako je $\mathbb{P}_X$ $(g.s.)$ aditivna, vrijedi i

$$\mathbb{P}_X(\tilde{B}_m) = \mathbb{P}_X(B_m) + \mathbb{P}_X(B_{m+1}) + ... + \mathbb{P}_X(B_{2^k}) = \mathbb{P}_X(B_m) \quad (g.s.).$$

(Ukoliko je $m = 2^k$, uzimamo $\tilde{B}_m = B_m$). Skupovi $B_1, ..., B_{m-1}, \tilde{B}_m$ sada čine izmjerivu particiju skupa $\mathbb{R}$, pri čemu vrijedi $\mathbb{P}_0(B_1), ..., \mathbb{P}_0(B_{m-1}), \mathbb{P}_0(\tilde{B}_m) > 0$. (Specijalno su $B_1, ..., B_{m-1}, \tilde{B}_m \neq \emptyset$). Ukoliko je ta particija jednočlana, odnosno $m = 1$ i $\tilde{B}_m = \mathbb{R}$, budući je $\mathbb{P}_X$ slučajna vjerojatnost, nužno je $\mathbb{P}_X(B_1) = \mathbb{P}_X(\tilde{B}_1) = 1 \ (g.s.)$. U suprotnom, točnije za $m \geq 2$, slijedeći definiciju 3.3.1, imamo

$$(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_m)) \sim \mathit{Dirichlet}\ (\alpha\mathbb{P}_0(B_1), ..., \alpha\mathbb{P}_0(B_m)). \tag{3.9}$$

(Umjesto $\tilde{B}_m$ pišemo B_m jer je $\mathbb{P}_X(\tilde{B}_m) = \mathbb{P}_X(B_m)$ $(g.s.)$ i $\mathbb{P}_0(\tilde{B}_m) = \mathbb{P}_0(B_m)$). Zbog načina na koji su skupovi $B_1, ..., B_m$ izabrani te koristeći $(g.s.)$ aditivnost od $\mathbb{P}_X$, distribuciju slučajnog vektora $(\mathbb{P}_X(A_1), ..., \mathbb{P}_X(A_k))$ dobivamo iz

$$\begin{aligned}(\mathbb{P}_X(A_1), ..., \mathbb{P}_X(A_k)) &= \left(\sum_{B_j \subseteq A_1} \mathbb{P}_X(B_j), ..., \sum_{B_j \subseteq A_k} \mathbb{P}_X(B_j) \right) = \\ &= \left(\sum_{\substack{B_j \subseteq A_1 \\ j \leq m}} \mathbb{P}_X(B_j), ..., \sum_{\substack{B_j \subseteq A_k \\ j \leq m}} \mathbb{P}_X(B_j) \right) \quad (g.s.).\end{aligned} \tag{3.10}$$

Nadalje, jasno je da za dane skupove $A_1, ..., A_k \in \mathcal{B}$ nije važno u kojem poretku promatramo pripadne $B_1, ..., B_{2^k}$. Stoga, ako je $(i_1, ..., i_k)$ proizvoljna permutacija od $(1, ..., k)$, direktno iz (3.10) dobivamo da je

$$F_{\mathbb{P}_X(A_{i_1}),...,\mathbb{P}_X(A_{i_k})}(y_{i_1}, ..., y_{i_k}) = F_{\mathbb{P}_X(A_1),...,\mathbb{P}_X(A_k)}(y_1, ..., y_k),$$

za svako $(y_1, ..., y_k) \in \mathbb{R}^k$. Drugim riječima, zadovoljen je uvjet *(1)* iz definicije 1.3.4.

Da bismo pokazali drugi Kolmogorovljev uvjet suglasnosti, pretpostavimo da su skupovima $A_1, ..., A_k$ dodani međusobno različiti skupovi $A_{k+1}, ..., A_r \in \mathcal{B}$ ($r \in \mathbb{N}$, $r > k$), svi različiti od $A_1, ..., A_k$. Analogno konstrukciji skupova $B_1, ..., B_{2^k}$, skupovima $A_1, ..., A_r$ pridružimo skupove $C_1, ..., C_{2^r}$. Neka za točno prvih m' tih skupova vrijedi $\mathbb{P}_0(C_j) > 0$ te neka je kao i ranije

$$(\mathbb{P}_X(C_1), ..., \mathbb{P}_X(C_{m'})) \sim \mathit{Dirichlet}\,(\alpha\mathbb{P}_0(C_1), ..., \alpha\mathbb{P}_0(C_{m'})). \tag{3.11}$$

Nije teško uočiti da particija $\{C_1, ..., C_{2^r}\}$ profinjuje particiju $\{B_1, ..., B_{2^k}\}$. Stoga imamo

$$\begin{aligned}(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_m)) &= \left(\sum_{\substack{C_j \subseteq B_1 \\ j \leq m'}} \mathbb{P}_X(C_j), ..., \sum_{\substack{C_j \subseteq B_m \\ j \leq m'}} \mathbb{P}_X(C_j) \right) \sim \left\{ \text{propozicija 3.2.6} \right\} \sim \\ &\sim \mathit{Dirichlet} \left(\sum_{\substack{C_j \subseteq B_1 \\ j \leq m'}} \alpha\mathbb{P}_0(C_j), ..., \sum_{\substack{C_j \subseteq B_m \\ j \leq m'}} \alpha\mathbb{P}_0(C_j) \right) \sim \\ &\sim \mathit{Dirichlet}\,(\alpha\mathbb{P}_0(B_1), ..., \alpha\mathbb{P}_0(B_m)).\end{aligned}$$

Stoga je (3.11) usklađeno sa (3.9). S druge strane, također vrijedi

$$(\mathbb{P}_X(A_1), ..., \mathbb{P}_X(A_k)) = \left(\sum_{\substack{C_j \subseteq A_1 \\ j \leq m'}} \mathbb{P}_X(C_j), ..., \sum_{\substack{C_j \subseteq A_k \\ j \leq m'}} \mathbb{P}_X(C_j) \right) = \left(\sum_{C_j \subseteq A_1} \mathbb{P}_X(C_j), ..., \sum_{C_j \subseteq A_k} \mathbb{P}_X(C_j) \right) =$$

$$= \left(\sum_{B_j \subseteq A_1} \mathbb{P}_X(B_j), ..., \sum_{B_j \subseteq A_k} \mathbb{P}_X(B_j) \right) = \left(\sum_{\substack{B_j \subseteq A_1 \\ j \leq m}} \mathbb{P}_X(B_j), ..., \sum_{\substack{B_j \subseteq A_k \\ j \leq m}} \mathbb{P}_X(B_j) \right).$$

Stoga je distribucija slučajnoga vektora $(\mathbb{P}_X(A_1), ..., \mathbb{P}_X(A_k))$, koja se dobije pomoću (3.10), usklađena s marginalnom distribucijom prvih k komponenata od $(\mathbb{P}_X(A_1), ..., \mathbb{P}_X(A_r))$, koja se dobije pomoću (3.11). Drugim riječima, zadovoljen je i uvjet *(2)* definicije 1.3.4.

Time smo opravdali egzistenciju Dirichletovog procesa zadanog definicijom 3.3.1.

3.4 Primjeri

Primjer 3.4.1. Neka je $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$ slučajna vjerojatnost, za neke $\alpha > 0$ i $\mathbb{P}_0 \in \mathcal{P}$. Neka je $B_1 \in \mathcal{B}$ takvo da je $\mathbb{P}_0(B_1) \in \langle 0, 1\rangle$ (tj. $\mathbb{P}_0(B_1)$, $\mathbb{P}_0(\mathbb{R} \setminus B_1) > 0$). Uzimajući $k = 2$ i $B_2 = \mathbb{R} \setminus B_1$, iz definicije Dirichletovog procesa dobivamo

$$\mathbb{P}_X(B_1) \sim Beta(\alpha\mathbb{P}_0(B_1), \alpha - \alpha\mathbb{P}_0(B_1)).$$

Iz primjera 1.1.9 znamo da je očekivanje takve slučajne varijable jednako

$$\mathbb{E}(\mathbb{P}_X(B_1)) = \frac{\alpha\mathbb{P}_0(B_1)}{\alpha\mathbb{P}_0(B_1) + \alpha - \alpha\mathbb{P}_0(B_1)} = \mathbb{P}_0(B_1).$$

Iako nismo formalno uveli pojam očekivanja slučajne vjerojatnosti, možemo reći da je očekivanje Dirichletovog procesa $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$ jednako baznoj vjerojatnosti $\mathbb{P}_0$, tj.

$$\mathbb{E}(\mathbb{P}_X) = \mathbb{P}_0.$$

Primjer 3.4.2. Promatramo podatke $x_1, ..., x_n \in \mathbb{R}$ $(n \in \mathbb{N})$, koji su realizacija slučajnog uzorka $X_1, ..., X_n$ iz nepoznate distribucije $\mathbb{P}_X$. Kao apriorni bayesovski model uzimamo da je $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$, gdje su $\alpha > 0$ i $\mathbb{P}_0 \in \mathcal{P}$ neka (konkretna, neslučajna) vjerojatnosna mjera na $(\mathbb{R}, \mathcal{B})$. Komentar na moguće izbore za α i $\mathbb{P}_0$ ostavljamo za kasnije. Dakle, $\mathbb{P}_X$ je slučajna vjerojatnost. Neka je sada $k \geq 2$ te $\{B_1, ..., B_k\}$ neka izmjeriva particija skupa $\mathbb{R}$, takva da je $\mathbb{P}_0(B_j) > 0$, za svako $j = 1, ..., k$. Tada je (po definiciji Dirichletovog procesa)

$$(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_k)) \sim Dirichlet(\alpha\mathbb{P}_0(B_1), ..., \alpha\mathbb{P}_0(B_k)).$$

Koristimo oznaku $\mathbb{X} = (X_1, ..., X_n)$. Iz primjera 3.2.4 slijedi

$$(\mathbb{P}_X(B_1), ..., \mathbb{P}_X(B_k))\,|\,\mathbb{X} \sim Dirichlet\left(\alpha\mathbb{P}_0(B_1) + \sum_{i=1}^{n} \delta_{x_i}(B_1), ..., \alpha\mathbb{P}_0(B_k) + \sum_{i=1}^{n} \delta_{x_i}(B_k)\right),$$

gdje je sa

$$\delta_{x_i}(B) = \begin{cases} 1, & x_i \in B, \\ 0, & x_i \notin B, \end{cases} \qquad B \in \mathcal{B}$$

označena mjera na $(\mathbb{R}, \mathcal{B})$ koncentrirana u točki x_i $(i = 1, ..., n)$. Dakle,

$$\sum_{i=1}^{n} \delta_{x_i}(B_j)$$

predstavlja broj podataka među $x_1, ..., x_n$ koji su upali u B_j, za $j = 1, ..., k$. Time smo dobili

$$\mathbb{P}_X\,|\,\mathbb{X} \sim DP\left(\alpha\mathbb{P}_0 + \sum_{i=1}^{n} \delta_{x_i}\right),$$

odnosno

$$\mathbb{P}_X\,|\,\mathbb{X} \sim DP\left((\alpha + n) \cdot \left(\frac{\alpha}{\alpha + n} \cdot \mathbb{P}_0 + \frac{n}{\alpha + n} \cdot \sum_{i=1}^{n} \frac{1}{n}\delta_{x_i}\right)\right). \tag{3.12}$$

Nije se teško uvjeriti da je

$$\frac{\alpha}{\alpha + n} \cdot \mathbb{P}_0 + \frac{n}{\alpha + n} \cdot \sum_{i=1}^{n} \frac{1}{n}\delta_{x_i}$$

zaista vjerojatnost na $(\mathbb{R}, \mathcal{B})$. Dakle, Dirichletov proces nasljeđuje poželjno svojstvo konjugiranosti od Dirichletove distribucije.
Promatrajmo sada slučaj $k = 2, B_1 = \langle -\infty, x]$, za neko $x \in \mathbb{R}$ i $B_2 = \mathbb{R} \setminus B_1$. Pod pretpostavkom da je $\mathbb{P}_0(\langle -\infty, x]) \in \langle 0, 1\rangle$, kao u primjeru 3.4.1 dobivamo

$$\mathbb{E}(\mathbb{P}_X(\langle -\infty, x])\,|\,\mathbb{X}) = \frac{\alpha}{\alpha + n} \cdot \mathbb{P}_0(\langle -\infty, x]) + \frac{n}{\alpha + n} \cdot \sum_{i=1}^{n} \frac{1}{n}\delta_{x_i}(\langle -\infty, x]). \tag{3.13}$$

Funkciju $x \mapsto \hat{F}_{X,B}(x) = \mathbb{E}\left((\mathbb{P}_X(\langle -\infty, x]) \,|\, \mathbb{X}\right)$ nazivat ćemo **bayesovskom procjenom funkcije distribucije**. Vidimo da je to težinska sredina funkcije distribucije pridružene početnoj baznoj vjerojatnosti $\mathbb{P}_0$

$$F_{\mathbb{P}_0}(x) = \mathbb{P}_0(\langle -\infty, x])$$

i uzoračke funkcije distribucije (eng. *empirical cumulative distribution function - ECDF*), koju koristimo i u frekvencionističkoj statistici

$$\hat{F}_X(x) = \sum_{i=1}^{n} \frac{1}{n} \delta_{x_i}(\langle -\infty, x]) \,.$$

Promatrajmo konkretne podatke $x_1, ..., x_n$, $n = 20$, koji su (nezavisno) simulirani iz eksponencijalne distribucije s očekivanjem 1. Podaci (sortirani) su dani u tablici 3.1. Za neprekidnu slučajnu varijablu X kažemo da ima **eksponencijalnu distribuciju s očekivanjem (parametrom)** $\lambda > 0$, ako joj je gustoća dana sa

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

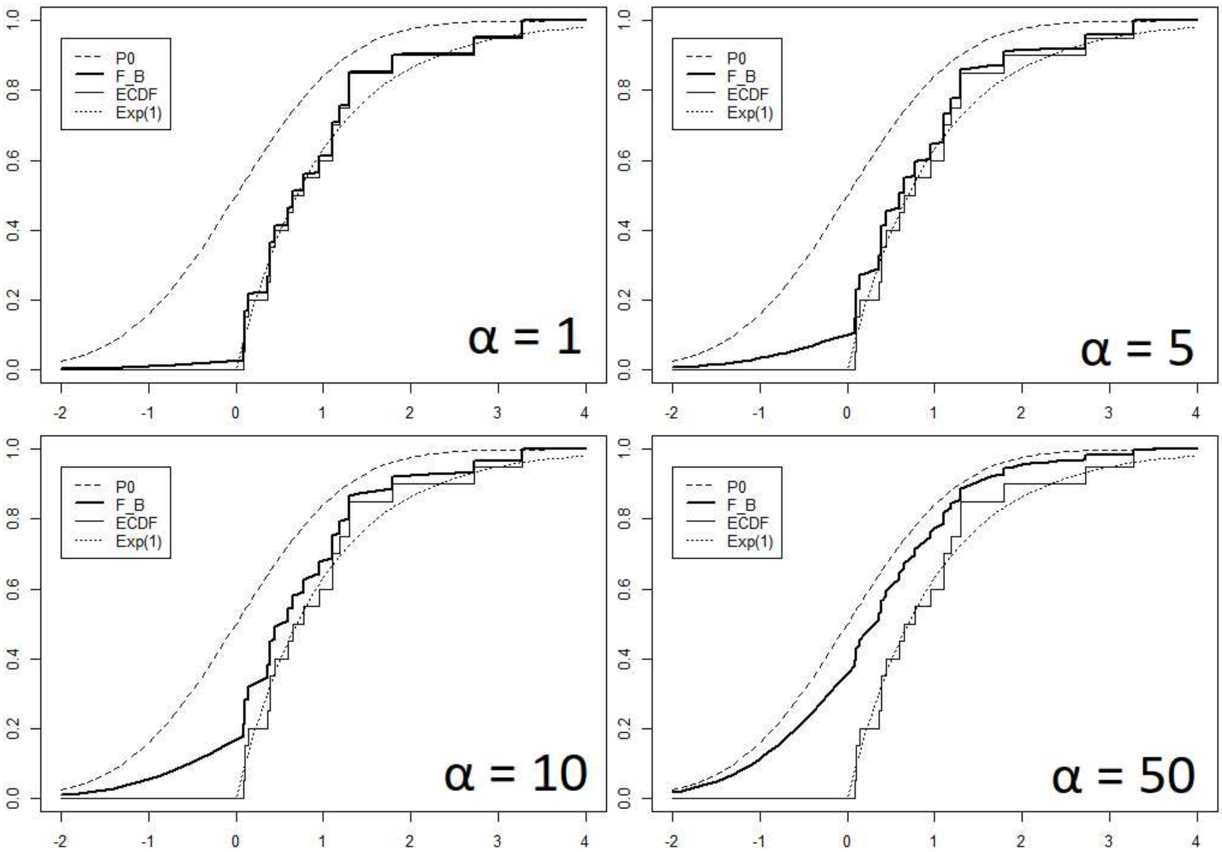
Pišemo $X \sim Exp(\lambda)$. Recimo da na početku vjerujemo da bi podaci trebali dolaziti iz

x_1	0.08615741	x_6	0.38934936	x_{11}	0.76981024	x_{16}	1.29038435
x_2	0.10351111	x_7	0.39266879	x_{12}	0.94759052	x_{17}	1.29570942
x_3	0.10624687	x_8	0.45033380	x_{13}	1.10441600	x_{18}	1.78639107
x_4	0.13467689	x_9	0.59301642	x_{14}	1.10451306	x_{19}	2.72291870
x_5	0.35575773	x_{10}	0.65172024	x_{15}	1.19199763	x_{20}	3.28302255

Tablica 3.1: Podaci za primjere 3.4.2 i 3.4.4

standardne normalne distribucije, tj. $N(0, 1)$ (v. primjer 2.3.2). Stoga za $\mathbb{P}_0$ uzimamo vjerojatnosnu mjeru induciranu tom distribucijom. Za skalirajući parametar uzimamo $\alpha = 1, 5, 10, 50$. Grafički uspoređujemo dobivene $\hat{F}_{X,B}$, $\hat{F}_X$ te funkcije distribucija $N(0, 1)$ i $Exp(1)$ razdiobe. Rezultati su prikazani na slici 3.1. Očekivano, za male vrijednosti parametra α dobivamo procjenu $\hat{F}_{X,B}$ koja je vrlo slična uzoračkoj funkciji distribucije $\hat{F}_X$, dok se za veće α približava krivulji funkcije distribucije od $\mathbb{P}_0$. Intuitivno je jasno da će se za $\alpha \ll n$, funkcija $\hat{F}_{X,B}$ ponašati približno jednako kao $\hat{F}_X$.

U primjeru 3.4.2 konstruirali smo bayesovsku procjenu funkcije distribucije, koja se temeljila na nekom polaznom modelu i opaženim podacima. Međutim, postavlja se pitanje možemo li model $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$ do kojeg smo došli (za neke nove $\alpha > 0$, $\mathbb{P}_0 \in \mathcal{P}$),



Slika 3.1: Rezultati procjene distribucije pomoću Dirichletovog procesa

iskoristiti bolje od toga da samo računamo očekivanja marginalnih jednodimenzionalnih distribucija. Također, uopće nije intuitivno jasno čemu bi bile jednake realizacije tog Dirichletovog procesa. Stoga je uobičajenije promatrati tzv. *proces razlomljenog štapa* (eng. *stick-breaking process*). U [6] je pokazano da $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$ ima sljedeću reprezentaciju

$$\mathbb{P}_X(\cdot) = \sum_{h=1}^{\infty} \pi_h \delta_{\theta_h}(\cdot), \quad \pi_h = v_h \cdot \prod_{j=1}^{h-1}(1 - v_j), \quad h \in \mathbb{N}, \tag{3.14}$$

gdje su

$$v_h \overset{iid}{\sim} Beta(1, \alpha), \quad \theta_h \overset{iid}{\sim} \mathbb{P}_0, \quad h \in \mathbb{N}.$$

Odavde se vidi da su realizacije od $\mathbb{P}_X$ $(g.s.)$ diskretne distribucije s nosačem $\{\theta_h : h \in \mathbb{N}\}$.

Svakom atomu θ_h pridružena je masa π_h, pri čemu vrijedi

$$\sum_{h=1}^{\infty} \pi_h = 1 \quad (g.s.).$$

Ukoliko su neka dva atoma θ_{h_1} i θ_{h_2} jednaka (ili više njih), njihove se mase zbrajaju. Slikovito objašnjenje načina na koji dolazimo do niza $(\pi_h)_{h\in\mathbb{N}}$ je sljedeći. Polazimo od štapa duljine 1 kojeg razlamamo na dva dijela duljina v_1 i $1 - v_1$. Potom drugi dio štapa, duljine $1 - v_1$, razlamamo na dva dijela novih duljina $v_2(1 - v_1)$ i $(1 - v_2)(1 - v_1)$. Postupak nastavljamo beskonačno mnogo puta. Budući $\pi_h \to 0$ za $h \to \infty$, u praksi se simuliraju $v_1, ..., v_N \sim Beta(1, \alpha)$, za neko dovoljno veliko $N \in \mathbb{N}$, a za $h > N$ se uzima $\pi_h = 0$.

Primjer 3.4.3. Neka je $\mathbb{P}_0$ vjerojatnosna mjera pridružena standardnoj normalnoj distribuciji (v. primjer 2.3.2). Promatrajmo slučajnu vjerojatnost $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$, za $\alpha = 0.5, 1, 5, 10$. Simulirajmo po 3 realizacije svakog takvog Dirichletovog procesa koristeći reprezentaciju (3.14). Uzimamo da su $\pi_h = 0$, za $h > 1000$. Funkcije gustoća dobivenih diskretnih distribucija prikazane su na slici 3.2. Vidimo da se za manje vrijednosti od α većina mase rezultirajućih distribucija koncentrira u manjem broju atoma nego za veće α.

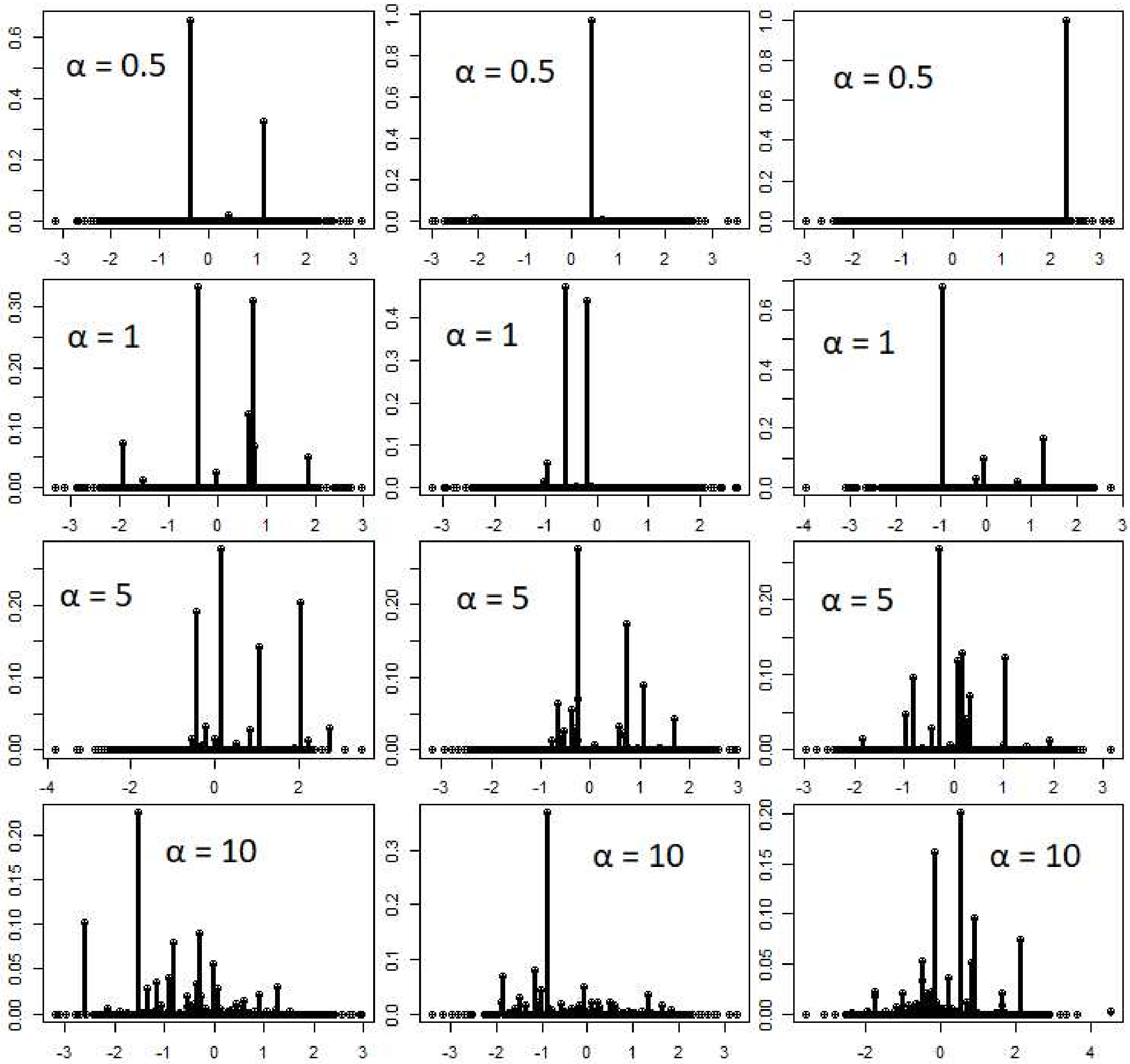
Primjer 3.4.4. Sada kada smo naučili uzorkovati iz Dirichletovog procesa, možemo se vratiti na primjer 3.4.2. Tamo smo došli do modela (3.12), gdje smo za $\mathbb{P}_0$ uzeli $\mathbb{P}_0 \sim N(0, 1)$, $\alpha = 1, 5, 10, 50$, a podaci $x_1, ..., x_{20}$ su bili dani tablicom 3.1. Za svaki α simulirajmo po 10 vjerojatnosnih distribucija $\mathbb{P}_{\alpha,1}, ..., \mathbb{P}_{\alpha,10}$ iz Dirichletovog procesa danog sa (3.12). Dakle, ovoga puta uzorkujemo $v_h \overset{iid}{\sim} Beta(1, \alpha + n)$, za $h = 1, ..., 1000$, dok za $h > 1000$, kao u primjeru 3.4.3, uzimamo $\pi_h = 0$. S druge pak strane,

$$\theta_h \overset{iid}{\sim} \frac{\alpha}{\alpha + n} \cdot \mathbb{P}_0 + \frac{n}{\alpha + n} \cdot \sum_{i=1}^{n} \frac{1}{n} \delta_{x_i}$$

uzorkujemo tako da s vjerojatnošću $\frac{\alpha}{\alpha+n}$ simuliramo θ_h iz $\mathbb{P}_0$, odnosno s vjerojatnošću $\frac{n}{\alpha+n}$ za θ_h odaberemo neku od točaka $x_1, ..., x_n$. Preciznije, možemo simulirati Bernoullijevu slučajnu varijablu $b_h \sim B(\frac{\alpha}{\alpha+n})$ (v. primjer 1.1.6). Zatim, u slučaju $b_h = 0$ uzorkujemo

$$\theta_h \sim \begin{pmatrix} x_1 & ... & x_{20} \\ 1/n & ... & 1/n \end{pmatrix}, \tag{3.15}$$

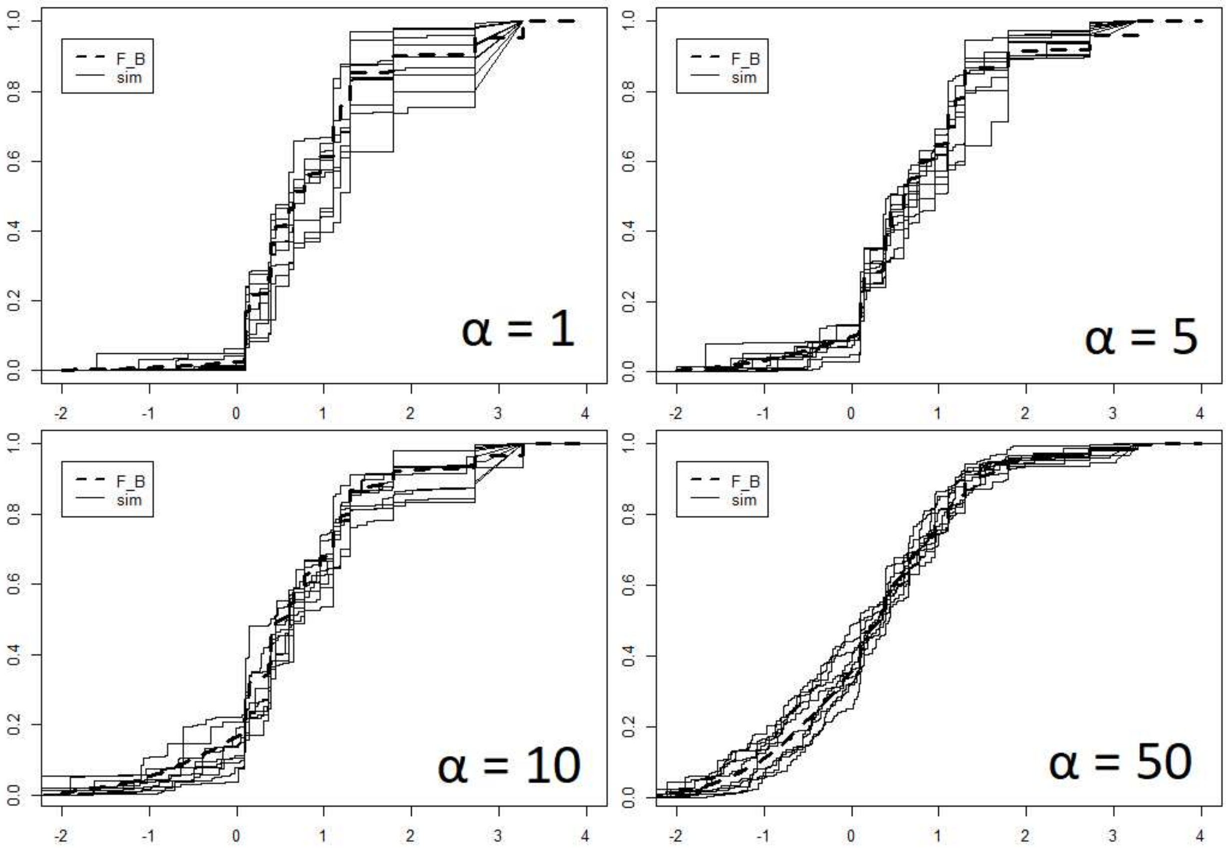
odnosno u slučaju $b_h = 1$ uzorkujemo $\theta_h \sim \mathbb{P}_0$. To činimo nezavisno za svako $h = 1, ..., 1000$. Na kraju koristimo (3.14). Ovoga puta crtamo funkcije distribucija dobivenih $\mathbb{P}_{\alpha,1}, ..., \mathbb{P}_{\alpha,10}$, koje na grafu odmah uspoređujemo s bayesovskom procjenom funkcije



Slika 3.2: Simulacije iz primjera 3.4.3

distribucije $\hat{F}_{X,B}$, danom sa (3.13). Rezultati su prikazani na slici 3.3. Vidimo da za veće α dobivamo $\mathbb{P}_{\alpha,1}, ..., \mathbb{P}_{\alpha,10}$ s manjim skokovima, što je i očekivano, jer za veće α θ_h češće uzorkujemo iz $\mathbb{P}_0$ nego iz (3.15). Ovdje ponovno treba napomenuti da $\mathbb{P}_{\alpha,1}, ..., \mathbb{P}_{\alpha,10}$ nismo simulirali iz $DP(\alpha\mathbb{P}_0)$, već iz (3.12).

Dakle, bez obzira na prirodu distribucije $\mathbb{P}_0$ od koje krećemo graditi model, iz (3.14)



Slika 3.3: Simulacije iz primjera 3.4.4

se vidi da će Dirichletov proces $\mathbb{P}_X \sim DP(\alpha\mathbb{P}_0)$ (kao i aposteriorni proces $\mathbb{P}_X \,|\, \mathbb{X}$) biti $(g.s.)$ diskretna vjerojatnosna distribucija. Stoga se Dirichletov proces uglavnom ne koristiti za modeliranje distribucija $\mathbb{P}_X$ za koje bismo htjeli da budu neprekidne. Ipak, Dirichletov se proces može koristiti u rješavanju raznih drugih problema (v. [2], 23.). Ovdje ćemo u kratkim crtama ilustrirati jednu moguću primjenu.

Primjer 3.4.5. Slično kao u poglavlju 2, promatramo statističko obilježje X čija funkcija gustoće ovisi o parametru Θ, koji je pak diskretna slučajna varijabla. Sa $\mathbb{P}_\Theta$ označimo vjerojatnosnu mjeru pridruženu distribuciji od Θ. Gustoću od X tada možemo zapisati kao

$$f_X(x) = \sum_\theta f_{X|\Theta}(x \,|\, \theta) \cdot \mathbb{P}_\Theta(\theta), \quad x \in \mathbb{R}, \tag{3.16}$$

pri čemu sumiramo po svim mogućim θ za koje je $\mathbb{P}_\Theta(\theta) \equiv \mathbb{P}_\Theta(\{\theta\}) > 0$. Model dan u (3.16) općenito nazivamo **mješoviti model** (eng. *mixed-model*). Primjerice, težina životinja je različito distribuirana od vrste do vrste. Također, vrijeme koje korisnik društvene mreže provede pregledavajući određeni sadržaj može ovisiti o dobnoj, spolnoj ili interesnoj skupini kojoj pripada. Dakle, slučajna varijabla Θ bilježi kojoj skupini pripada opažanje X, dok $f_{X|\Theta}(\cdot\,|\,\theta)$ opisuje distribuciju od X unutar (konkretne) skupine θ. Posebno zanimljivi slučajevi mješovitih modela su oni u kojima postoji beskonačno (tj. prebrojivo) mnogo mogućih skupina ili je njihov broj konačan, ali vrlo velik i ne može se procijeniti gornja granica. Upravo je to slučaj i kada govorimo o broju životinjskih vrsta na Zemlji ili profilima ljudi koji se koriste društvenim mrežama. Budući je Θ diskretna slučajna varijabla, za konstrukciju modela distribucije od Θ može se koristiti upravo Dirichletov proces $\mathbb{P}_\Theta \sim DP(\alpha\mathbb{P}_0)$. Više o primjenama Dirichletovog procesa u ovakvim problemima, može se pročitati u [2], 23.3.

Bibliografija

[1] T. S. Ferguson, *A Bayesian Analysis of Some Nonparametric Problems*, The Annals of Statistics **1** (1973), br. 2, 209–230.

[2] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari i D. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, 2013, `http://www.stat.columbia.edu/~gelman/book/BDA3.pdf`, (7. rujan 2022.).

[3] V. Jović, *Poisson-Dirichletova razdioba*, Magistarski rad, 2015, `https://urn.nsk.hr/urn:nbn:hr:217:786739`, (7. rujan 2022.).

[4] R. Mrazović, *Mjera i integral, skripta Matematičkog odsjeka Prirodoslovno-matematičkog fakulteta u Zagrebu*, `https://web.math.pmf.unizg.hr/nastava/mii/files/mii-predavanja.pdf`, (7. rujan 2022.).

[5] N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, 2002.

[6] J. Sethuraman, *A constructive definition of Dirichlet priors*, Statistica Sinica (1994), br. 2, 639–650, `https://www2.stat.duke.edu/~scs/Courses/Stat376/Papers/DirichletProc/DirichletSummary.pdf`, (7. rujan 2022.).

Sažetak

U ovom radu uvodimo pojam *Dirichletovog procesa* u kontekstu *bayesovske statistike*. Navedeni su osnovni pojmovi i rezultati teorije vjerojatnosti koji će se koristiti u nastavku rada, a neke pomoćne tvrdnje su i dokazane. Posebno, definiran je pojam slučajnog procesa te je opisan način na koji se on zadaje pomoću suglasne familije konačno-dimenzionalnih distribucija. Zatim je definirana uvjetna funkcija gustoće, pokazana je Bayesova formula te su objašnjeni pojmovi *apriorne* i *aposteriorne* gustoće. Kroz nekoliko je primjera opisano zaključivanje u tzv. *parametarskoj bayesovskoj statistici*. Konačno, u kontekstu tzv. *neparametarske bayesovske statistike* opisan je pojam slučajne vjerojatnosne mjere, specijalnog slučaja slučajnoga procesa. Uvedena je Dirichletova distribucija, a Dirichletov proces definiran je kao ona slučajna vjerojatnost čije su konačno-dimenzionalne distribucije upravo Dirichletove. Definicija je opravdana provjerom Kolmogorovljevih uvjeta suglasnosti. Teorija je zaokružena primjerima i malom simulacijskom studijom.

Summary

In this thesis, we introduce the term *Dirichlet process* in the context of *Bayesian statistics*. The basic terms and results of the probability theory that will be used in the rest of the paper are listed, and some auxiliary propositions are also proven. In particular, the concept of a stochastic process is defined and the way in which it is assigned using a consistent collection of finite-dimensional distributions is described. Then the conditional probability density function is defined, the Bayes' formula is shown and the terms *prior* and *posterior* density functions are explained. Several examples describe the conclusion in the so-called *parametric Bayesian statistics*. Finally, in the context of the so-called *non-parametric Bayesian statistics* the concept of a random probability measure, a special case of a stochastic process, is described. The Dirichlet distribution is introduced, and the Dirichlet process is defined as that random probability measure whose finite-dimensional distributions are exactly Dirichlet. The definition is justified by checking the conditions required by the Kolmogorov extension theorem. The theory is rounded off with examples and a small simulation study.

Životopis

Rođen sam u Zagrebu 30. travnja 1998. godine. Osnovnu školu kralja Tomislava završio sam 2013. godine kao nagrađeni učenik generacije. Potom upisujem XV. gimnaziju u Zagrebu. Tijekom školovanja sudjelovao sam na državnim natjecanjima iz matematike (2013., 2016.), fizike (2013., 2014., 2015.) i logike (2016., 2017.) Preddiplomski sveučilišni studij matematike na zagrebačkom Prirodoslovno-matematičkom fakultetu upisujem 2017. godine, a diplomski sveučilišni studij matematičke statistike 2020. Tijekom sve tri godine preddiplomskoga studija ostvarivao sam pravo na tzv. STEM-stipendiju Ministarstva znanosti i obrazovanja, a od 2020. godine sam stipendist Privredne banke Zagreb.