# Koncentracijske nejednakosti i primjene

Jukić, Marko

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Marko Jukić

# KONCENTRACIJSKE NEJEDNAKOSTI I PRIMJENE

Diplomski rad

Voditelj rada:
prof. dr. sc. Bojan Basrak

Zagreb, 2021.

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. ____________________________

# Sadržaj

# Uvod

Zakoni velikih brojeva nam govore kako se sume nezavisnih, jednako distribuiranih slučajnih varijabli s velikom vjerojatnošću koncentriraju oko svog očekivanja. Takvo ponašanje dijele i mnogi drugi slučajni objekti. Dok su zakoni velikih brojeva po prirodi asimptotski, od većeg interesa će nam biti neasimptotske koncentracijske nejednakosti. Koncentracijske nejednakosti poprimaju razne oblike, poput ograda na očekivanje, varijancu ili rep distribucije slučajne varijable. Pojačani interes posljednjih desetljeća za ovaj tip vjerojatnosnih rezultata dolazi uglavnom iz područja statističkog učenja. U ovom radu prikazat ćemo neke metode dokazivanja koncentracijskih nejednakosti i njihove primjene u teoriji statističkog učenja.

U prvom poglavlju uvest ćemo neke temeljne pojmove iz Vapnik–Chervonenkisove teorije statističkog učenja s kojima ćemo se susretati u kasnijim poglavljima, poput VC dimenzije i Rademacherove kompleksnosti.

U drugom poglavlju pokazat ćemo neke osnovne koncentracijske nejednakosti, a fokus će biti na sumama nezavisnih slučajnih varijabli. Koristeći Chernoffljevu metodu pokazat ćemo eksponencijalne ograde na rep distribucije poput Hoeffdingove i Bernsteinove nejednakosti, a zatim ćemo pokazati da slične nejednakosti vrijede za subgaussovske i subeksponencijalne varijable.

U trećem poglavlju bavit ćemo se koncentracijskim nejednakostima za generalne funkcije nezavisnih slučajnih varijabli. Glavna ideja će biti da fluktuacije funkcije ograničimo preko fluktuacija u svakoj varijabli. Glavni rezultati će biti Efron-Steinova nejednakost i McDiarmidova nejednakost. Rezultati iz ovog dijela posebno su lako primjenjivi na funkcijama sa svojstvom omeđenih razlika i samoomeđujućim svojstvom, kakve često susrećemo u teoriji statističkog učenja.

U četvrtom poglavlju prikazat ćemo Logaritamske Sobolevljeve nejednakosti, na kojima se temelji metoda entropije. Ovom metodom mogu se dobiti eksponencijalne ograde na rep distribucije za samoomeđujuće funckije.

U petom poglavlju bavit ćemo se supremumom slučajnog procesa. Glavni rezultat će biti Dudleyeva nejednakost, koja daje ogradu na očekivanje preko metričke entropije. Pokazat ćemo i kako iz Dudleyeve nejednakosti slijede neke od temeljnih tvrdnji iz teorije statističkog učenja.

Rad se temelji na tekstu *Concentration-of-measure inequalities* Gábora Lugosija [4] te 2. i 8. poglavlju knjige *High-dimensional probability: An introduction with applications in data science* Romana Vershynina [7].

# Poglavlje 1

# Kratak uvod u teoriju statističkog učenja

U ovom poglavlju definirat ćemo neke važne pojmove iz teorije statističkog učenja, s kojima ćemo se susretati u ostalim poglavljima. Nećemo previše ulaziti u dubinu, već ćemo samo dati intuiciju iza pojedinih pojmova.

Teorija statističkog učenja bavi se pronalaskom optimalne funkcije veze između dvije slučajne varijable $X$ i $Y$ koje poprimaju vrijednosti u skupovima $\mathcal{X}$ i $\mathcal{Y}$ redom, i takve da vrijednosti $(X, Y)$ dolaze iz neke vjerojatnosne distribucije na produktnom izmjerivom prostoru $\mathcal{X} \times \mathcal{Y}$. Drugim riječima, tražimo funkciju koja predviđa vrijednosti varijable $\mathcal{Y}$ pomoću varijable $\mathcal{X}$. Tipični primjeri su prepoznavanje znakova na slici ili procjena cijene nekretnine pomoću raznih podataka o nekretnini.

Optimalnu funkciju veze biramo pomoću funkcije gubitka, kako zovemo izmjerive funkcije $l : \mathcal{Y}^2 \mapsto [0, \infty\rangle$. Želimo da optimalna funkcija veze ima što manji rizik.

**Definicija 1.1.** *Neka je $l : \mathcal{Y}^2 \mapsto [0, \infty\rangle$ funkcija gubitka. Funkciju rizika $R(f)$ izmjerive funkcije $f : \mathcal{X} \to \mathcal{Y}$ definiramo kao*

$$R(f) := \mathbb{E}[l(f(X), Y)] = \int_{\mathcal{X}\times\mathcal{Y}} l(f(x), y) d\mathbb{P}_{(X,Y)}(x, y).$$

Kada je $\mathcal{Y} = \mathbb{R}$, česti odabir funkcije gubitka je kvadratna pogreška

$$l(\hat{y}, y) = (\hat{y} - y)^2.$$

U slučaju klasifikacije, odnosno kada je $|Y| < \infty$, često koristimo 0-1 gubitak

$$l(\hat{y}, y) = \begin{cases} 1 & \hat{y} = y \\ 0 & \hat{y} \neq y \end{cases}.$$

Ako nam je distribucija od $(X, Y)$ poznata, često eksplicitno možemo odrediti funkciju $\phi$ koja minimizira rizik

$$\phi = \operatorname*{arg\,min}_{f \colon \mathcal{X} \to \mathcal{Y} \text{ izmjeriva}} R(f).$$

Ispod navodimo dva klasična primjera u kojima možemo odrediti optimalnu funkciju veze.

(i) Neka je $\mathcal{X} = \mathbb{R}^p$, $\mathcal{Y} = \mathbb{R}$ i neka je $l$ kvadratna pogreška. Onda je

$$\phi(x) = \mathbb{E}[Y \mid X = x]$$

i $\phi$ zovemo regresijska funkcija.

(ii) Neka je $|Y| < \infty$ i neka je $l$ funkcija 0-1 gubitka. Onda je

$$\phi(x) = \operatorname*{arg\,max}_{y \in \mathcal{Y}} \mathbb{P}(Y = y | X = x)$$

i $\phi$ zovemo Bayesov klasifikator.

Međutim, funkcija distribucije od $(X, Y)$ obično nije poznata, pa optimalnu funkciju biramo na temelju konačnog uzorka nezavisnih kopija od $(X, Y)$. To možemo napraviti tako da umjesto rizika minimiziramo empirijski rizik, koji bi trebao aproksimirati rizik odabrane funkcije.

**Definicija 1.2.** *Neka su $(X_1, Y_1), \dots, (X_n, Y_n)$ nezavisne kopije* [1] *od $(X, Y)$. Empirijsku funkciju rizika $R_n(f)$ izmjerive funkcije $f \colon \mathcal{X} \to \mathcal{Y}$ definiramo kao*

$$R_n(f) = \frac{1}{n} \sum_{i=1}^{n} l(f(X_i), Y_i).$$

Odmah vidimo jedan problem s minimizacijom empirijskog rizika za dani uzorak: može postojati beskonačno mnogo izmjerivih funkcija koje minimiziraju empirijski rizik. Nadalje, rješavanje problema minimizacije može biti prekomplicirano, kao i računanje vrijednosti dobivene optimalne funkcije veze. Zato ćemo se u izboru optimalne funkcije veze unaprijed ograničiti na neku familiju izmjerivih funkcija $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$. Upravo je odabir familije $\mathcal{F}$ ključan problem u teoriji statističkog učenja.

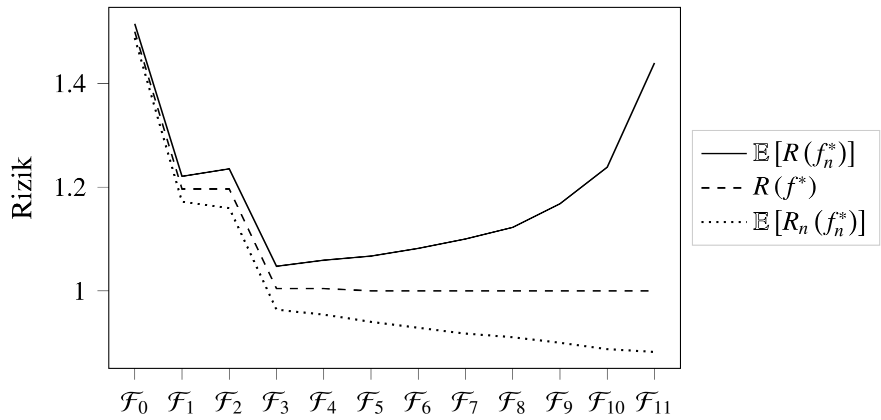
Označimo sa $f^* \in \mathcal{F}$ funkciju koja minimizira rizik, a sa $f_n^* \in \mathcal{F}$ slučajnu funkciju koja minimizira empirijski rizik. Tada rizik od $f_n^*$ možemo zapisati kao

$$R(f_n^*) = R(f^*) + (R(f_n^*) - R(f^*)).$$

[1] Kada uvodimo slučajne varijable koje su nezavisne kopije drugih varijabli, bez posebnog naglašavanja podrazumijevamo da su nezavisne i od svih drugih varijabli, u ovom slučaju su to $(X, Y)$.

Jasno je da ćemo s većim familijama postići manji rizik $R(f^*)$. Da bi minimizacija empirijskog rizika dala dobar rezultat, važno je odabrati familiju $\mathcal{F}$ takvu da je rizik $R(f_n^*)$ s velikom vjerojatnosti blizu $R(f^*)$, neovisno o distribuciji $(X, Y)$. Bez dodatnih pretpostavki ne može se puno reći o veličini $R(f^*)$, pa ćemo se zato fokusirati na veličinu $R(f_n^*) - R(f^*)$.

Da bismo bolje razumjeli problem odabira familije $\mathcal{F}$, analizirajmo rizik familija različitih veličina na jednom primjeru. Neka je $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, neka $X$ ima uniformnu distribuciju $X \sim U(0, 2\pi)$ i neka je $Y = \sin(X) + \varepsilon$ gdje je $\varepsilon \sim N(0, 1)$ slučajni šum nezavisan od $X$. Neka je veličina uzorka $n = 100$ i uzmimo da je $l$ kvadratna pogreška. Na Slici 1.1 je dana usporedba rizika za familije $\mathcal{F}_0, \ldots, \mathcal{F}_{11}$, gdje je $F_i$ familija realnih polinoma stupnja $\leq i$.




Slika 1.1: Usporedba rizika za familije različite veličine. Vrijednosti na grafu nisu egzaktne, već su izračunate Monte Carlo simulacijom.

Jasno je da će veće familije dati manji empirijski rizik. Ako je familija prevelika, kod minimizacije empirijskog rizika doći će do prevelikog prilagođavanja slučajnom šumu $\varepsilon$. Kao rezultat će rizik često biti velik iako je empirijski rizik malen. Taj koncept je poznat pod imenom *overfitting*, i to je razlog velikog rizika familije $\mathcal{F}_{11}$. Obratno, ako je familija premala, očekivano je da daje veći empirijski rizik, ali vidimo da empirijski rizik za manje familije puno bolje aproksimira pravi rizik. U ovom primjeru najmanji rizik daje $\mathcal{F}_3$, što nam govori da je optimalan odabir familija koja nije niti prevelika niti premala i kao takva daje mali empirijski rizik koji dobro aproksimira pravi rizik.

Rizik od $f_n^*$ također možemo zapisati na sljedeći način

$$R(f_n^*) = R_n(f_n^*) + (R(f_n^*) - R_n(f_n^*)).$$

Vidjeli smo da je za prekompleksne familije empirijski rizik $R_n(f_n^*)$ mali, ali on loše aproksimira rizik pa je vrijednost $R(f_n^*) - R_n(f_n^*)$ velika. Obratno, za premalu familiju, na primjer konačnu $|\mathcal{F}| < \infty$, slabi zakon velikih brojeva nam govori da vrijedi

$$\left|R(f_n^*) - R_n(f_n^*)\right| \leq \sum_{f \in \mathcal{F}} \underbrace{|R(f) - R_n(f)|}_{\xrightarrow[n\to\infty]{\mathbb{P}} 0} \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

pa znamo da će za dovoljno veliki uzorak empirijski rizik biti dobra aproksimacija rizika, odnosno vrijednost $R(f_n^*) - R_n(f_n^*)$ će biti mala. Problem je što ovako odabrana konačna familija $\mathcal{F}$ može biti prejednostavna i može imati velik empirijski rizik $R_n(f_n^*)$. Dodatan faktor kod izbora familije će biti veličina uzorka $n$, jer će kompleksnije familije zahtijevati veći uzorak da bi empirijski rizik dobro aproksimirao pravi rizik.

U ostatku rada često ćemo se, radi jednostavnosti, ograničiti na slučaj binarne klasifikacije, što znači da ćemo raditi sa $Y = \{0, 1\}$ funkcijom 0-1 gubitka. U razmatranju kompleksnosti različitih familija $\mathcal{F}$ pomoći će nam sljedeći pojmovi.

**Definicija 1.3.** *Neka je $\mathcal{X}$ skup i $\mathcal{F} \subseteq \{0, 1\}^{\mathcal{X}}$. Kažemo da se skup $S \subseteq \mathcal{X}$ može rastaviti sa $\mathcal{F}$ ako je svaka funkcija $g\colon S \to \{0, 1\}$ restrikcija neke funkcije $f \in \mathcal{F}$ na $S$.*

Primijetimo da ako se neki skup $S$ može rastaviti sa $\mathcal{F}$, onda se može rastaviti i sa svakom većom familijom $\mathcal{F}' \supseteq \mathcal{F}$. To nam govori da će se s većim i kompleksnijim familijama $\mathcal{F}$ moći rastaviti veći skupovi. Također, ako se sa $\mathcal{F}$ može rastaviti neki skup $S$, onda se može rastaviti i svaki $S' \subseteq S$. Tako dolazimo do sljedeće mjere kompleksnosti familije $\mathcal{F}$.

**Definicija 1.4** (VC dimenzija)**.** *Vapnik–Chervonenkisova dimenzija od $\mathcal{F} \subseteq \{0, 1\}^{\mathcal{X}}$, oznaka $\mathrm{vc}(\mathcal{F})$, je kardinalnost najvećeg $S \subseteq \mathcal{X}$ koji se može rastaviti sa $\mathcal{F}$*

$$\mathrm{vc}(\mathcal{F}) = \sup_{\substack{S \subseteq \mathcal{X}\colon S \text{ se može} \\ \text{rastaviti sa } \mathcal{F}}} |S|.$$

Navedimo nekoliko primjera familija $\mathcal{F}$ i njihove $VC$ dimenzije.

**Primjer 1.5** (Intervali u $\mathbb{R}$)**.** *Neka je $\mathcal{X} := \mathbb{R}$ i $\mathcal{F}$ familija karakterističnih funkcija zatvorenih intervala u $\mathbb{R}$*

$$\mathcal{F} := \{\mathbb{1}_{[a,b]} : a, b \in \mathbb{R}, a \leq b\}.$$

*Tvrdimo da je $\mathrm{vc}(\mathcal{F}) = 2$. Provjerimo da se $S = \{3, 5\}$ može rastaviti sa $\mathcal{F}$. Zaista, ako definiramo funkcije $f_{0,0} = \mathbb{1}_{\{0\}}, f_{0,1} = \mathbb{1}_{\{5\}}, f_{1,0} = \mathbb{1}_{\{3\}}, f_{1,1} = \mathbb{1}_{[3,5]}$, za svaku $g\colon S \to \{0, 1\}$ vrijedi*

$$g = f_{g(3),g(5)}\big|_S\,.$$

*Dakle, $\mathrm{vc}(\mathcal{F}) \geq 2$. Pokažimo sada $\mathrm{vc}(\mathcal{F}) < 3$, odnosno da se sa $\mathcal{F}$ ne može rastaviti nijedan skup veličine 3. Neka je $S = \{x, y, z\}$, za $x < y < z$ i neka je $g\colon S \to \{0, 1\}$ takva da je $g(x) = g(z) = 1$ i $g(y) = 0$. Kada bi postojala funkcija $f = \mathbb{1}_{[a,b]} \in \mathcal{F}$ takva da je $g = f|_S$, vrijedilo bi $x, z \in [a, b]$ i $y \notin [a, b]$, a to nije moguće.*

Ključna će biti razlika između familija konačne i beskonačne *VC* dimenzije.

**Primjer 1.6** (Beskonačna VC dimenzija)**.** *Neka je $\mathcal{X} = \mathbb{R}$ i neka je $\mathcal{F}$ familija svih Borelovih funkcija*

$$\mathcal{F} = \{f : \mathbb{R} \to \{0,1\} : f \text{ Borelova}\} = \{\mathbb{1}_B : B \subseteq \mathbb{R} \text{ Borelov}\}.$$

*Tada je očito* $\mathrm{vc}(\mathcal{F}) = \infty$*, jer se sa $\mathcal{F}$ mogu rastaviti proizvoljno veliki skupovi $S \subseteq \mathbb{R}$. Isto vrijedi i za familiju karakterističnih funkcija konačnih skupova*

$$\mathcal{F} = \{\mathbb{1}_B : B \subseteq \mathbb{R} \text{ Borelov}, |B| < \infty\}.$$

Jasno je da izmjerive funkcije $f : \mathcal{X} \to \{0,1\}$ možemo reprezentirati s izmjerivim skupovima $f^{-1}(\{1\})$. Slično možemo dati i definiciju VC dimenzije za familije $\mathcal{A}$ izmjerivih podskupova od $\mathcal{X}$, takvu da je $\mathrm{vc}(\mathcal{A}) = \mathrm{vc}(\{\mathbb{1}_A : A \in \mathcal{A}\})$. Navedimo još nekoliko primjera VC dimenzije preko takvih familija $\mathcal{A}$

- Neka je $\mathcal{X} = \mathbb{R}^d$ i neka je $\mathcal{A}$ familija svih poluravnina $\left\{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{x}, \boldsymbol{\alpha} \rangle + \beta \geq 0\right\}$, za $\boldsymbol{\alpha} \in \mathbb{R}^d, \beta \in \mathbb{R}$. Tada je $\mathrm{vc}(\mathcal{A}) = d + 1$.
- Neka je $\mathcal{X} = \mathbb{R}^d$ i neka je $\mathcal{A}$ familija svih kugli $\left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \boldsymbol{\alpha}\|_2 \leq \beta\right\}$, za $\boldsymbol{\alpha} \in \mathbb{R}^d, \beta \geq 0$. Tada je $\mathrm{vc}(\mathcal{A}) = d + 1$.
- Neka je $\mathcal{X} = \mathbb{R}^d$, $d \geq 2$ i neka je $\mathcal{A}$ familija svih zatvorenih konveksnih skupova $A \subseteq \mathbb{R}^d$. Tada je $\mathrm{vc}(\mathcal{A}) = \infty$.
- Iz prethodnih primjera mogli bismo naslutiti da je VC dimenzija povezana s brojem parametara u parametrizaciji familije $\mathcal{A}$, odnosno $\mathcal{F}$. To nije nužno slučaj, jer familiju $\mathcal{A}$ svih skupova oblika $\{x \in \mathbb{R} : \sin(tx) \geq 0\} \subseteq \mathbb{R}$ za $t \in \mathbb{R}$ možemo parametrizirati samo jednim parametrom $t$, a vrijedi $\mathrm{vc}(\mathcal{A}) = \infty$.

Još jedan način promatranja kompleksnosti familije je Rademacherova kompleksnost, koja mjeri koliko se familija $\mathcal{F}$ može prilagoditi slučajnom šumu.

**Definicija 1.7** (Rademacherova distribucija)**.** *Kažemo da je slučajna varijabla $X$ Rademacherova ako vrijedi*

$$X \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \iff \mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

**Definicija 1.8** (Rademacherova kompleksnost)**.** *Neka su $X_1, \ldots, X_n$ nezavisne i jednako distribuirane slučajne varijable i neka su $\sigma_1, \ldots, \sigma_n$ nezavisne Rademacherove slučajne* *varijable, nezavisne od $X_1, \dots, X_n$. Rademacherovu kompleksnost familije Borelovih funkcija $\mathcal{F}$ definiramo kao*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i f(X_i)\right].$$

*Ponekad se Rademacherova kompleksnost definira i s apsolutnom vrijednosti*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \sigma_i f(X_i)\right|\right].$$

Intuitivno, izraz unutar supremuma podsjeća na uzoračku kovarijancu/korelaciju slučajnog vektora $(f(X_1), \dots, f(X_n))$ i slučajnog šuma $(\sigma_1, \dots, \sigma_n)$. Zbog nezavisnosti je $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \sigma_i f(X_i)\right] = 0$ za svaku $f \in \mathcal{F}$, pa u slučaju velike Rademacherove kompleksnosti familije $\mathcal{F}$ možemo naslutiti da se familija prilagođava slučajnom šumu. To je obično znak da je familija $\mathcal{F}$ prekompleksna i da empirijski rizik neće biti dobra aproksimacija rizika odabrane funkcije veze.

**Definicija 1.9** (Uniformna Glivenko-Cantellijeva klasa)**.** *Kažemo da je familija Borelovih funkcija $\mathcal{F}$ s vrijednostima u nekom skupu $\mathcal{X}$ uniformna Glivenko-Cantellijeva klasa ako za svaki $\varepsilon > 0$ vrijedi*

$$\lim_{n\to\infty} \sup_{\mathbb{P}_X} \mathbb{P}\left(\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right| > \varepsilon\right) = 0, \tag{1.1}$$

*gdje je prvi supremum po svim vjerojatnosnim mjerama $\mathbb{P}_X$ na $\mathcal{X}$, a slučajne varijable $X, X_1, \dots, X_n$ su nezavisne jednako distribuirane s razdiobom vrijednosti $\mathbb{P}_X$. Gornji limes možemo preciznije zapisati kao*

$$\lim_{n\to\infty} \sup_{\mathbb{P}_X} (\mathbb{P}_X)^n\left(\left\{(x_1, \dots, x_n) \in \mathcal{X}^n \colon \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \int_{\mathcal{X}} f d\mathbb{P}_X\right| > \varepsilon\right\}\right) = 0.$$

Neka je $\mathcal{F}$ neka familija Borelovih funkcija $f\colon \mathcal{X} \to \mathcal{Y}$. Definirajmo familiju $\mathcal{G}$ sa

$$\mathcal{G} = \{(x, y) \mapsto l(f(x), y) \colon f \in \mathcal{F}\}.$$

Po definiciji, familija $\mathcal{G}$ će biti uniformna Glivenko-Cantellijeva klasa ako i samo ako za svaki $\varepsilon > 0$ vrijedi

$$\lim_{n\to\infty} \sup_{\mathbb{P}_{X,Y}} \mathbb{P}\left(\sup_{f\in\mathcal{F}} |R_n(f) - R(f)| > \varepsilon\right) = 0.$$

Ovdje sada vidimo jasnu povezanost s problemom odabira familije $\mathcal{F}$. Ako je $\mathcal{G}$ uniformna Glivenko-Cantellijeva klasa, za dovoljno velike uzorke će empirijski rizik biti dobra aproksimacija rizika, neovisno o distribuciji iz koje dolaze podaci. Obratno, ako $\mathcal{G}$ nije uniformna Glivenko-Cantellijeva klasa, empirijski rizik neće dobro aproksimirati rizik za neke distribucije podataka.

# Poglavlje 2

# Osnovne nejednakosti

U ovom poglavlju bavit ćemo se ogradama na rep distribucije. Ograde se razlikuju u pretpostavkama na slučajne varijable, a jače pretpostavke naravno daju i bolju ogradu. Najjednostavnija ograda na rep je Markovljeva nejednakost, koja uz pretpostavku konačnog očekivanja slučajne varijable daje ogradu na rep s redom veličine $t^{-1}$.

**Teorem 2.1** (Markovljeva nejednakost)**.** *Za nenegativnu slučajnu varijablu $X$ i $t > 0$ vrijedi*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

*Dokaz.* Vidi [5, Korolar 10.3]. □

Čebiševljeva nejednakost zahtijeva konačnu varijancu, odnosno konačan drugi moment i daje ogradu na rep s redom veličine $t^{-2}$.

**Teorem 2.2** (Čebiševljeva nejednakost)**.** *Neka je $X$ slučajna varijabla i $\mathbb{E}[|X|] < \infty$. Za $t > 0$ vrijedi*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

*Dokaz.* Neka je $t > 0$. Koristeći Markovljevu nejednakost dobivamo

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}\Big((X - \mathbb{E}[X])^2 \geq t^2\Big) \overset{2.1}{\leq} \frac{\mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big]}{t^2} = \frac{\mathrm{Var}(X)}{t^2}.$$ □

Sljedeći teorem daje jednostrano poboljšanje Čebiševljeve nejednakosti.

**Teorem 2.3** (Čebišev-Cantellijeva nejednakost)**.** *Uz iste pretpostavke za $t > 0$ vrijedi*

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + t^2}.$$

*Dokaz.* Vidi [4, Theorem 1]. □

Primijetimo da, kao u dokazu Čebiševljeve nejednakosti, za svaki $q > 0$ vrijedi

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^q]}{t^q}.$$

Ako su svi momenti konačni, gornju ogradu bismo mogli optimizirati po $q$ i tako dobiti

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \inf_{q>0} \frac{\mathbb{E}[|X - \mathbb{E}[X]|^q]}{t^q}. \tag{2.1}$$

Iako optimizacija po $q$ daje bolju ogradu, Čebiševljeva nejednakost je često puno praktičnija. Na primjer, u slučaju sume nezavisnih slučajnih varijabli $X = \sum_{i=1}^n X_i$ vrijedi aditivnost varijance $\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(X_i)$, što znači da je dovoljno analizirati varijance pojedinih varijabli $X_i$. Optimizacija ograde po parametru pokazat će se kao korisna ideja i upravo to je glavna ideja Chernoffljeve metode.

**Teorem 2.4** (Chernoffljeva nejednakost)**.** *Za slučajnu varijablu $X$, $t \in \mathbb{R}$ vrijedi*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}\left[e^{sX}\right]}{e^{st}}, \quad \textit{za svaki } s \geq 0 \tag{2.2}$$

$$\mathbb{P}(X \leq -t) \leq \frac{\mathbb{E}\left[e^{sX}\right]}{e^{-st}}, \quad \textit{za svaki } s \leq 0. \tag{2.3}$$

*Dokaz.* Primijetimo da je funkcija $x \mapsto e^{sx}$ monotona. Nejednakost će slijediti iz Markovljeve nejednakosti. Za $s \geq 0$ vrijedi

$$\mathbb{P}(X \geq t) \leq \mathbb{P}\left(e^{sX} \geq e^{st}\right) \overset{2.1}{\leq} \frac{\mathbb{E}\left[e^{sX}\right]}{e^{st}}.$$

Za $s \leq 0$ vrijedi

$$\mathbb{P}(X \leq -t) \leq \mathbb{P}\left(e^{sX} \geq e^{-st}\right) \overset{2.1}{\leq} \frac{\mathbb{E}\left[e^{sX}\right]}{e^{-st}}.$$ □

Primijetimo da se u Chernoffljevoj nejednakosti pojavljuje funkcija izvodnica momenata $M_X(s) := \mathbb{E}\left[e^{sX}\right]$ od $X$. Ako je $M_X(s) < \infty$ za neki $s \neq 0$, dolazimo do eksponencijalne ograde za rep vjerojatnosti, što je znatno poboljšanje u odnosu na Čebiševljevu nejednakost. Znamo da ako je $M_X(s)$ konačna na nekoj okolini nule, onda su svi momenti konačni, što znači da je konačnost funkcije izvodnice momenata jača pretpostavka od konačnosti momenata.

Ideja *Chernoffljeve metode* računanja ograda repova je da ogradimo desnu stranu Chernoffljeve nejednakosti i novu ogradu optimiziramo po varijabli $s$. Demonstrirajmo metodu na jednom primjeru.

**Primjer 2.5.** *Neka je $X \sim N(0, \sigma^2)$. Tada za svaki $s \in \mathbb{R}$ vrijedi*

$$\mathbb{E}\Big[e^{sX}\Big] = e^{\frac{\sigma^2 s^2}{2}}.$$

*Zato za $s, t \geq 0$ imamo*

$$\mathbb{P}(X \geq t) \overset{(2.2)}{\leq} \frac{\mathbb{E}\Big[e^{sX}\Big]}{e^{st}} = \exp\left(\frac{\sigma^2 s^2}{2} - st\right).$$

*Funkcija u eksponentu je kvadratna u $s$ i postiže minimum u $s^* = \frac{t}{\sigma^2} \geq 0$. Uvrštavanjem $s^*$ za svaki $t \geq 0$ dobivamo optimalnu Chernoffljevu ogradu za desni rep*

$$\mathbb{P}(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

*Za usporedbu, Čebiševljeva nejednakost u ovom slučaju daje*

$$\mathbb{P}(X \geq t) \leq \frac{\sigma^2}{t^2},$$

*pa vidimo da Chernoffljeva metoda daje puno bolju, eskponencijalnu ogradu.*

Kao što smo vidjeli u prethodnom primjeru, Chernoffljeva metoda obično daje puno bolje ograde od Čebiševljeve nejednakosti. Pokažimo da je Chernoffljeva ograda iz prethodnog primjera "točna" do na faktor $\frac{1}{t}$.

**Propozicija 2.6.** *Neka je $X \sim N(0, 1)$. Tada za svaki $t > 0$ vrijedi*

$$\frac{t}{t^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \leq \mathbb{P}(X \geq t) \leq \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

*Dokaz.* Pokažimo gornju ogradu

$$\begin{aligned}
\mathbb{P}(X \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}} dx \leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{x}{t} e^{-\frac{x^2}{2}} dx \\
&= \begin{bmatrix} y = \frac{x^2}{2} \\ dy = x dx \end{bmatrix} = \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} \int_{\frac{t^2}{2}}^\infty e^{-y} dy \\
&= \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.
\end{aligned}$$

Pokažimo donju ogradu. Neka je $g : [0, \infty\rangle \to \mathbb{R}$ definirana sa

$$g(t) = \mathbb{P}(X \geq t) - \frac{t}{t^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

Očito je $g(0) > 0$ i $\lim_{t\to\infty} g(t) = 0$. Lako dobijemo

$$g'(t) = -\frac{1}{\sqrt{2\pi}} \cdot \frac{2e^{-\frac{t^2}{2}}}{(t^2+1)^2} < 0,$$

iz čega vidimo da je $g$ strogo padajuća, pa mora biti $g > 0$. □

Vidjeli smo da se Chernoffljevom metodom mogu dobiti jako oštre ograde. Primijetimo da za $X \geq 0$ i $s > 0$ vrijedi

$$\frac{\mathbb{E}\left[e^{sX}\right]}{e^{st}} = \frac{1}{e^{st}} \sum_{n=0}^{\infty} \mathbb{E}\left[\frac{(sX)^n}{n!}\right] = \frac{1}{e^{st}} \sum_{n=0}^{\infty} \frac{(st)^n}{n!} \frac{\mathbb{E}[X^n]}{t^n} \geq \left(\inf_{n\in\mathbb{N}_0} \frac{\mathbb{E}[X^n]}{t^n}\right) \cdot \frac{1}{e^{st}} \sum_{n=0}^{\infty} \frac{(st)^n}{n!} = \inf_{n\in\mathbb{N}_0} \frac{\mathbb{E}[X^n]}{t^n},$$

pa računanjem infimuma po svim $s > 0$ dobivamo

$$\inf_{n\in\mathbb{N}_0} \frac{\mathbb{E}[X^n]}{t^n} \leq \inf_{s>0} \frac{\mathbb{E}\left[e^{sX}\right]}{e^{st}} = \inf_{s>0} \frac{M_X(s)}{e^{st}}.$$

To znači da su moment ograde, kao u (2.1), uvijek bolje od Chernoffljevih. Chernoffljeva ograda je ipak obično puno praktičnija, jer u slučaju sume nezavisnih varijabli $X = \sum_{i=1}^n X_i$ imamo $M_X(s) = \prod_{i=1}^n M_{X_i}(s)$, što znači da se onda možemo baviti ogradama na pojedine $M_{X_i}(s)$. U nastavku ćemo vidjeti da je takav pristup jako koristan.

## 2.1 Sume nezavisnih slučajnih varijabli

U ovom odjeljku bavit ćemo se koncentracijskim nejednakostima za sume nezavisnih slučajnih varijabli. Takve su česta pojava u teoriji statističkog učenja, kao na primjer empirijski rizik neke funkcije $f$ na uzorku $(X_i, Y_i)_{i=1}^n$.

Usporedimo prvo što o koncentraciji sume nezavisnih varijabli govore Centralnim granični teorem i Čebiševljeva nejednakost. Neka su $X_1, \ldots, X_n$ nezavisne, jednako distribuirane slučajne varijable s varijancom $\sigma^2$. Označimo $S_n = \sum_{i=1}^n X_i$ i neka je $t > 0$. Centralni granični teorem daje

$$\mathbb{P}\Big(S_n - \mathbb{E}[S_n] \geq t\sqrt{n}\Big) = \mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}\sigma} \geq \frac{t}{\sigma}\right) \approx 1 - \Phi\left(\frac{t}{\sigma}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \tag{2.4}$$

dok iz Čebiševljeve nejednakosti slijedi

$$\mathbb{P}\Big(|S_n - \mathbb{E}[S_n]| \geq t\sqrt{n}\Big) \leq \frac{\mathrm{Var}(S_n)}{nt^2} = \frac{\sigma^2}{t^2}. \tag{2.5}$$

Obje nejednakosti govore da su "tipične" devijacije slučajne varijable $S_n$ od njenog očekivanja reda veličine $\sqrt{n}$. Čebiševljeva nejednakost daje nejednakost koja nije asimptotska i vrijedi čak i kad $X_i$ nisu jednako distribuirane. Centralni granični teorem daje puno bolju eksponencijalnu ogradu na vjerojatnost, ali ta ograda je asimptotska i može imati veliku grešku. Sljedeći primjer govori o grešci u aproksimaciji centralnim graničnim teoremom, koja može pokvariti točnost asimptotske ograde (2.4).

**Primjer 2.7** (Greška aproksimacije centralnim graničnim teoremom)**.** *Neka su $X_1, \ldots, X_n$ nezavisne jednako distribuirane slučajne varijable s konačnim trećim momentom. Neka je $S_n = \sum_{i=1}^n X_i$ i $Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$. Označimo sa $F_n$ i $\Phi$ funkcije distribucije od $Z_n$ i standardne normalne distribucije, redom. Berry-Esseen centralni granični teorem kaže da za svaki $t \in \mathbb{R}$ vrijedi* [1]

$$|F_n(t) - \Phi(t)| \lesssim \frac{1}{\sqrt{n}} \mathbb{E}\left[\left|\frac{X_1 - \mathbb{E}[X_1]}{\sqrt{\mathrm{Var}(X_1)}}\right|^3\right].$$

*Dakle, greška u uniformnoj aproksimaciji najviše je reda $\frac{1}{\sqrt{n}}$. Neka je $X_i \sim B\left(1, \frac{1}{2}\right)$. Koristeći Stirlingovu aproksimaciju $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ dobivamo*

$$|F_n(0) - \Phi(0)| = \frac{\mathbb{P}(Z_{2n} = 0)}{2} = \frac{\mathbb{P}(S_{2n} = n)}{2} = 2^{-2n-1}\frac{(2n)!}{(n!)^2} \sim 2^{-2n-1}\frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}}{\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right)^2} = \frac{1}{2\sqrt{\pi n}}.$$

[1] Oznaku $\lesssim$ koristimo za nejednakosti koje vrijede ako se desna strana pomnoži nekom apsolutnom konstantom $C > 0$. Za konstantu kažemo da je apsolutna ako ne ovisi o drugim varijablama, u ovom slučaju $C$ ne ovisi o $n$ i $X_1, \ldots, X_n$, dok god su pretpostavke zadovoljene.

*Zaključujemo da red veličine greške u uniformnoj aproksimaciji općenito ne može biti manji od $\frac{1}{\sqrt{n}}$, što je puno veći red veličine od same ograde* (2.4).

Iz gornjeg primjera vidimo da potrebu za ne-asimptotskim rezultatima, jer asimptotski mogu imati veliku grešku. U ovom odjeljku pokazat ćemo (ne-asimptotske) koncentracijske nejednakosti koje uz pretpostavku ograničenosti varijabli $X_i$ daju eksponencijalnu ogradu jako sličnu (2.4). Prva je Hoeffdingova nejednakost, za koju će nam trebati Hoeffdingova lema. Općenitija, uvjetna Hoeffdingova lema koju sada dokazujemo bit će nam također korisna kasnije.

**Lema 2.8** (Uvjetna Hoeffdingova lema)**.** *Neka je $X$ slučajna varijabla na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ i neka je $\mathcal{G}$ $\sigma$-podalgebra od $\mathcal{F}$. Ako postoje $\mathcal{G}$-izmjerive slučajne varijable $A, B$ takve da je $A \leq X \leq B$ g.s., onda za sve $s \in \mathbb{R}$ g.s. vrijedi*

$$\mathbb{E}\Big[e^{s(X-\mathbb{E}[X \mid \mathcal{G}])} \;\Big|\; \mathcal{G}\Big] \leq \exp\left(\frac{s^2(B-A)^2}{8}\right).$$

*Dokaz.* Jednakosti i nejednakosti u dokazu vrijede gotovo sigurno, ako nije drugačije naglašeno. Bez smanjenja općenitosti možemo pretpostaviti $\mathbb{E}[X \mid \mathcal{G}] = 0$, jer inače definiramo slučajnu varijablu $X' = X - \mathbb{E}[X \mid \mathcal{G}]$, za koju vrijedi $\mathbb{E}[X' \mid \mathcal{G}] = 0$ i $A - \mathbb{E}[X \mid \mathcal{G}] \leq X' \leq A - \mathbb{E}[X \mid \mathcal{G}]$. Ako pokažemo lemu za $X'$, vrijedit će

$$\begin{aligned}\mathbb{E}\Big[e^{s(X-\mathbb{E}[X \mid \mathcal{G}])} \;\Big|\; \mathcal{G}\Big] &= \mathbb{E}\Big[e^{sX'} \;\Big|\; \mathcal{G}\Big] \\ &\leq \exp\left(\frac{s^2((B-\mathbb{E}[X \mid \mathcal{G}]) - (A - \mathbb{E}[X \mid \mathcal{G}]))^2}{8}\right) \\ &= \exp\left(\frac{s^2(B-A)^2}{8}\right).\end{aligned}$$

Sada zbog $\mathbb{E}[X \mid \mathcal{G}] = 0$ računanjem uvjetnog očekivanja dobivamo $A = \mathbb{E}[A \mid \mathcal{G}] \leq 0 \leq \mathbb{E}[B \mid \mathcal{G}] = B$. Na događaju $\{A \geq B\}$ je $X = A = B = 0$, pa umjesto nejednakosti vrijedi jednakost

$$\mathbb{E}\Big[e^{sX} \;\Big|\; \mathcal{G}\Big] \cdot \mathbb{1}_{\{A \geq B\}} = \mathbb{E}\Big[e^{sX} \cdot \mathbb{1}_{\{A \geq B\}} \;\Big|\; \mathcal{G}\Big] = \mathbb{1}_{\{A \geq B\}} = \exp\left(\frac{s^2(B-A)^2}{8}\right) \cdot \mathbb{1}_{\{A \geq B\}}.$$

Zato je dovoljno nejednakost pokazati na događaju $\{A < B\}$. Definirajmo slučajne varijable $P = \frac{-A}{B-A}\mathbb{1}_{\{A<B\}}$ i $U = s(B - A)$. Eksponencijalna funkcija je konveksna, pa po Jensenovoj nejednakosti na $\{A < B\}$ po točkama vrijedi

$$\exp(sX) = \exp\left(\frac{B-X}{B-A} \cdot sA + \frac{X-A}{B-A} \cdot sB\right) \leq \frac{B-X}{B-A}e^{sA} + \frac{X-A}{B-A}e^{sB}.$$

Računanjem uvjetnog očekivanja dobivamo

$$\begin{aligned}
\mathbb{E}\Big[e^{sX} \cdot \mathbb{1}_{\{A<B\}} \;\Big|\; \mathcal{G}\Big] &\leq \left(\frac{B - \mathbb{E}[X \mid \mathcal{G}]}{B-A} e^{sA} + \frac{\mathbb{E}[X \mid \mathcal{G}] - A}{B-A} e^{sB}\right) \cdot \mathbb{1}_{\{A<B\}} \\
&= \left(\frac{B}{B-A} e^{sA} + \frac{-A}{B-A} e^{sB}\right) \cdot \mathbb{1}_{\{A<B\}} \\
&= (1 - P + Pe^{U}) e^{-PU} \cdot \mathbb{1}_{\{A<B\}} \\
&\leq \exp\left(\frac{U^2}{8}\right) \cdot \mathbb{1}_{\{A<B\}} \\
&= \exp\left(\frac{s^2(B-A)^2}{8}\right) \cdot \mathbb{1}_{\{A<B\}},
\end{aligned}$$

gdje smo u zadnjoj nejednakosti iskoristili da za $p, u \geq 0$ vrijedi

$$-pu + \log(1 - p + pe^{u}) \leq \frac{1}{8} u^2.$$

Pokažimo još tu nejednakost. Označimo $f(u) = -pu + \log(1 - p + pe^{u})$. Za funkciju $f$ vrijedi

$$\begin{aligned}
f(0) &= 0 \\
f'(0) &= -p + \frac{p}{p + (1-p)e^{-u}}\bigg|_{u=0} = 0 \\
f''(u) &= \frac{p(1-p)e^{-u}}{(p + \underbrace{(1-p)e^{-u}}_{q})^2} = \frac{pq}{(p+q)^2} \leq \frac{1}{4}.
\end{aligned}$$

Po Taylorovom teoremu, za svaki $u \in \mathbb{R}$ postoji $u^*$ između 0 i $u$ takav da vrijedi

$$f(u) = f(0) + f'(0)u + \frac{f''(u^*)}{2} u^2 \leq \frac{1}{8} u^2. \qquad \square$$

**Lema 2.9** (Hoeffdingova lema)**.** *Neka je $X$ slučajna varijabla takva da je $X \in [a, b]$ g.s. Tada za sve sve $s \in \mathbb{R}$ vrijedi*

$$\mathbb{E}\Big[e^{s(X - \mathbb{E}[X])}\Big] \leq \exp\left(\frac{s^2(b-a)^2}{8}\right).$$

*Dokaz.* Slijedi iz uvjetne Hoeffdingove leme uvrštavanjem $\mathcal{G} = \{\emptyset, \Omega\}$, $A = a$, $B = b$

$$\mathbb{E}\Big[e^{s(X - \mathbb{E}[X])}\Big] = \mathbb{E}\Big[e^{s(X - \mathbb{E}[X \mid \mathcal{G}])} \;\Big|\; \mathcal{G}\Big] \leq \exp\left(\frac{s^2(B-A)^2}{8}\right) = \exp\left(\frac{s^2(b-a)^2}{8}\right). \qquad \square$$

**Teorem 2.10** (Hoeffdingova nejednakost)**.** *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable takve da je $X_i \in [a_i, b_i]$ g.s. Ako označimo $S_n = X_1 + \cdots + X_n$, tada za svaki $t \geq 0$ vrijedi*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{2.6}$$

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \leq -t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{2.7}$$

*Dokaz.* Za dokaz nejednakosti koristit ćemo Chernoffljevu ogradu i Hoeffdingovu lemu. Za $s \geq 0$ vrijedi

$$\begin{aligned}
\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) &\overset{(2.2)}{\leq} e^{-st}\mathbb{E}\left[e^{s(S_n - \mathbb{E}[S_n])}\right] \\
&= e^{-st}\mathbb{E}\left[\prod_{i=1}^n e^{s(X_i - \mathbb{E}[X_i])}\right] \\
&= e^{-st}\prod_{i=1}^n \mathbb{E}\left[e^{s(X_i - \mathbb{E}[X_i])}\right] \\
&\overset{2.9}{\leq} e^{-st}\prod_{i=1}^n \exp\left(\frac{s^2(b_i - a_i)^2}{8}\right) \\
&= \exp\left(\frac{s^2\sum_{i=1}^n (b_i - a_i)^2}{8} - st\right).
\end{aligned}$$

Izraz u eksponentu je kvadratni u $s$, pa ga možemo minimizirati uvrštavanjem $s^* = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2} \geq 0$ i tako dobijemo upravo nejednakost (2.6). Nejednakost (2.7) dokazuje se tako da na početku umjesto (2.2) iskoristimo (2.3) i dalje nastavimo analogno. □

Nastavimo usporedbu s Centralnim graničnim teoremom. Pretpostavimo sada dodatno da su $X_i$ ograničene, odnosno da je $|X_i - \mathbb{E}[X_i]| \leq c$, za svaki $i$. Primijetimo da je tada $\sigma \leq c$ i Hoeffdingova nejednakost daje

$$\mathbb{P}\left(|S_n - \mathbb{E}[S_n]| \geq t\sqrt{n}\right) \leq 2\exp\left(-\frac{t^2}{2c^2}\right). \tag{2.8}$$

Dobili smo nejednakost istog oblika kao (2.4), ali sa $c$ umjesto $\sigma$. Primijetimo da za $X_i \sim B\left(1, \frac{1}{2}\right)$ vrijedi $\sigma = \frac{1}{2} = c$, pa u tom slučaju (2.4) i (2.8) daju istu ogradu. Ovdje opet vidimo veliku prednost ne-asimptotske nejednakosti, a to je da ne ovisi o distribuciji $X_i$.

Mana Hoeffdingove nejednakosti je da ne koristi informaciju o varijanci, pa za $\sigma \ll c$ Hoeffdingova nejednakost daje puno slabiju ogradu od Centralnog graničnog teorema. Jedan primjer takve situacije je $X_i \sim B(1, p)$, $p \ll \frac{1}{2}$. U situacijama kada znamo nešto o varijancama slučajnih varijabli $X_i$, Bennetova i Bernsteinova nejednakost mogu dati poboljšanje u odnosu na Hoeffdingovu.

**Teorem 2.11** (Bennetova nejednakost)**.** *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable takve da za sve vrijedi $\mathbb{E}[X_i] = 0$ i $|X_i| \leq c$ g.s. Neka je $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[X_i^2\big] > 0$. Tada za sve $t \geq 0$ vrijedi*

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{n\sigma^2}{c^2} h\left(\frac{ct}{n\sigma^2}\right)\right).$$

*Gdje je $h \colon [0, +\infty\rangle \to \mathbb{R}$ zadana sa*

$$h(u) = (1+u)\log(1+u) - u.$$

*Dokaz.* Koristeći Chernoffljevu ogradu i nezavisnost, za $s \geq 0$ dobivamo

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \overset{(2.2)}{\leq} e^{-st}\mathbb{E}\left[e^{s\sum_{i=1}^n X_i}\right] = e^{-st}\prod_{i=1}^n \mathbb{E}\left[e^{sX_i}\right].$$

Računamo

$$\begin{aligned}
\mathbb{E}\left[e^{sX_i}\right] &= \mathbb{E}\left[1 + \sum_{k=1}^\infty \frac{(sX_i)^k}{k!}\right] \\
&= 1 + \sum_{k=2}^\infty \frac{s^k \mathbb{E}\left[X_i^k\right]}{k!} && \text{LTDK}, \mathbb{E}[X_i] = 0 \\
&\leq 1 + \sum_{k=2}^\infty \frac{s^k c^{k-2}\mathbb{E}\left[X_i^2\right]}{k!} && X_i^k \leq c^{k-2}X_i^2 \text{ za } k \geq 2 \\
&= 1 + \frac{\mathbb{E}\left[X_i^2\right]}{c^2}\sum_{k=2}^\infty \frac{(sc)^k}{k!} \\
&= 1 + \frac{\mathbb{E}\left[X_i^2\right]}{c^2}(e^{sc} - 1 - sc) \\
&\leq \exp\left(\frac{\mathbb{E}\left[X_i^2\right](e^{sc} - 1 - sc)}{c^2}\right). && 1 + x \leq e^x
\end{aligned}$$

Iskoristimo dobivene nejednakosti

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq e^{-st}\prod_{i=1}^n \mathbb{E}\left[e^{sX_i}\right] \leq \exp\left(\frac{n\sigma^2(e^{sc} - 1 - sc)}{c^2} - st\right).$$

Izračunajmo optimalni $s \geq 0$

$$0 = \frac{\partial}{\partial s}\left(\frac{n\sigma^2(e^{sc} - 1 - sc)}{c^2} - st\right) = \frac{n\sigma^2 e^{cs} - 1}{c} - t \implies s^* = \frac{1}{c}\log\left(1 + \frac{ct}{n\sigma^2}\right) \geq 0.$$

Uvrštavanjem $s^*$ dobivamo traženu nejednakost. □

**Teorem 2.12** (Bernsteinova nejednakost)**.** *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable takve da za sve vrijedi $\mathbb{E}[X_i] = 0$ i $|X_i| \le c$ g.s. Neka je $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] > 0$. Tada za sve $t \ge 0$ vrijedi*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \ge t\right) \le \exp\left(-\frac{nt^2}{2\sigma^2 + \frac{2}{3}ct}\right).$$

*Dokaz.* Nejednakost slijedi iz Bennetove nejednakosti nakon zamjene funkcije $h(u)$ manjom funkcijom $g(u) := \frac{u^2}{2+\frac{2}{3}u}$. Pokažimo da za sve $u \ge 0$ vrijedi

$$h(u) \ge g(u) \iff \underbrace{(1+u)\log(1+u) - u - \frac{u^2}{2+\frac{2}{3}u}}_{f(u)} \ge 0.$$

Po Taylorovom teoremu vrijedi da za svaki $u \ge 0$ postoji $u^* \in [0, u]$ takav da vrijedi

$$f(u) = f(0) + f'(0)u + \frac{f''(u^*)}{2}u^2.$$

Očito je $f(0) = 0$. Također vrijedi

$$\begin{aligned} f'(0) &= \log(1+u) - \frac{3u(u+6)}{2(u+3)^2}\bigg|_{u=0} = 0 \\ f''(u) &= \frac{u^2(u+9)}{(u+1)(u+3)^2} \ge 0 \text{ za } u \ge 0. \end{aligned}$$

Zaključujemo:

$$f(u) = \frac{f''(u^*)}{2}u^2 \ge 0.$$ □

Stavimo li izraz unutar vjerojatnosti u isti oblik kao u Centralnom graničnom teoremu, Bernsteinova nejednakost daje

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\sqrt{n}\right) \le \exp\left(-\frac{t^2}{2\sigma^2 + \frac{2}{3}c\frac{t}{\sqrt{n}}}\right).$$

Neka je $C > 0$ proizvoljna konstanta. Slijedi da za $t \le C\sqrt{n}$ vrijedi

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\sqrt{n}\right) \le \exp\left(-\frac{t^2}{2\sigma^2 + \frac{2}{3}cC}\right),$$

što opet govori da su tipične devijacije reda veličine $\sqrt{n}$, a ograda na vjerojatnost je istog oblika kao kod Centralnog graničnog teorema. Također, interval vrijednosti $t \in \langle 0, C\sqrt{n}]$

za koje vrijedi gornja nejednakost raste kako se $n$ povećava, baš kao što aproksimativna ograda (2.4) istog oblika postaje sve točnija za $n \to \infty$. Za $t \geq C\sqrt{n}$ imamo

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\sqrt{n}\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\frac{t}{C\sqrt{n}} + \frac{2}{3}c\frac{t}{\sqrt{n}}}\right) = \exp\left(-\frac{t\sqrt{n}}{\frac{2\sigma^2}{C} + \frac{2}{3}c}\right),$$

pa smo za velike devijacije dobili drukčiju ogradu, koja umjesto $t^2$ ima $t$ u eksponentu i podsjeća na rep eksponencijalne distribucije. U sljedećim odjeljcima ćemo generalizirati Hoeffdingovu i Bernsteinovu nejednakost za šire familije distribucija, koje ćemo zvati subgaussovske i subeskponencijalne.

## 2.2 Subgaussovske slučajne varijable

U prethodnom odjeljku smo vidjeli da je ograničenost varijabli $X_1, \ldots, X_n$ dovoljan uvjet da bi vrijedila nejednakost oblika

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\sqrt{n}\right) \leq 2e^{-ct^2},$$

gdje je $c > 0$ konstanta koja može ovisiti o $X_1, \ldots, X_n$. Pitamo se kako možemo karakterizirati varijable $X_i$ za koje vrijede nejednakosti gornjeg oblika. Uzmemo li $n = 1$, odmah vidimo da mora vrijediti ograda na rep

$$\mathbb{P}(|X_1| \geq t) \leq 2e^{-ct^2}.$$

Slučajne varijable koje zadovoljavaju gornje svojstvo zvat ćemo subgaussovske, a vidjet ćemo da za takve također vrijedi Hoeffdingova nejednakost.

**Definicija 2.13** (Subgaussovska slučajna varijabla)**.** *Kažemo da je slučajna varijabla $X$ subgaussovska ako zadovoljava neko od sljedećih ekvivalentnih svojstava. Pozitivne konstante $K_i$ koje se pojavljuju u donjim svojstvima razlikuju se najviše za faktor koji je apsolutna konstanta, što kraće možemo zapisati kao $K_i \lesssim K_j$, za svaki $i$, $j$.*

*(i) Repovi od $X$ zadovoljavaju*

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right) \quad \textit{za svaki } t \geq 0.$$

*(ii) Momenti od $X$ zadovoljavaju*

$$\|X\|_{L^p} = \mathbb{E}[|X|^p]^{\frac{1}{p}} \leq K_2\sqrt{p} \quad \textit{za svaki } p \geq 1.$$

*(iii) Funkcija izvodnica momenata od $X^2$ zadovoljava*

$$\mathbb{E}\left[e^{sX^2}\right] \leq e^{K_3^2 s} \quad \text{za svaki } s \in \left[0, \frac{1}{\sqrt{K_3}}\right].$$

*(iv) Funkcija izvodnica momenata od $X^2$ zadovoljava*

$$\mathbb{E}\left[\exp\left(\frac{X^2}{K_4^2}\right)\right] \leq 2.$$

*Ako je $\mathbb{E}[X] = 0$, onda su svojstva (i)-(iv) ekvivalentna sljedećem svojstvu.*

*(v) Funkcija izvodnica momenata od X zadovoljava*

$$\mathbb{E}\left[e^{sX}\right] \leq e^{K_5^2 s^2} \quad \text{za svaki } s \in \mathbb{R}$$

*Za dokaz ekvivalencije svojstava (i)-(v) vidi [7, Proposition 2.5.2].*

Korisno je promatrati najmanje konstante koje zadovoljavaju gornja svojstva, pa tako dolazimo do definicije subgasussovske norme.

**Definicija 2.14** (Subgaussovska norma)**.** *Neka je X subgaussovska slučajna varijabla. Subgaussovsku normu definiramo kao najmanji $K_4$ iz svojstva (iv) iz prethodne definicije*

$$\|X\|_{\psi_2} = \inf\left\{s > 0 \colon \mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right] \leq 2\right\}.$$

**Napomena 2.15.** *Svojstva (i)-(v) iz Definicije 2.13 možemo zapisati preko subgaussovske norme. Na primjer, vrijedi*

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{Ct^2}{\|X\|_{\psi_2}^2}\right) \quad \text{za svaki } t \geq 0,$$

*gdje je $C > 0$ apsolutna konstanta. Također vrijedi da je $\|X\|_{\psi_2}$, do na apsolutno konstantan faktor, najmanji broj za kojeg vrijedi gornja nejednakost.*

Jasno je da je slučajna varijabla subgaussovska ako i samo ako ima konačnu subgaussovsku normu. Već smo vidjeli da ograda za rep iz svojstva (i) vrijedi za centriranu normalnu distribuciju, a nije teško provjeriti da su i sve ograničene slučajne varijable subgaussovske.

**Primjer 2.16.** *Navedimo nekoliko primjera subgaussovske norme.*

*(a) Neka je $X$ slučajna varijabla s Rademacherovom distribucijom. Ekvivalentno je*

$$\mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right] \leq 2 \iff \exp\left(\frac{1}{s^2}\right) \leq 2 \iff s \geq \frac{1}{\sqrt{\ln(2)}}.$$

*Slijedi da je $\|X\|_{\psi_2} = \frac{1}{\sqrt{\ln(2)}}$.*

*(b) Slično kao u (a), za ograničenu slučajnu varijablu $X$ vrijedi*

$$\|X\|_{\psi_2} \leq \frac{1}{\sqrt{\ln(2)}}\|X\|_\infty. \tag{2.9}$$

*(c) Neka je $X \sim N(0, \sigma^2)$. Koristeći ogradu iz Primjera 2.5 za $X$ i $-X$, lako dobivamo*

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right).$$

*Sada iz Napomene 2.15 slijedi $\|X\|_{\psi_2} \lesssim \sigma = \|X\|_{L^2}$. Preciznije, znamo da je $\left(\frac{X}{\sigma}\right)^2 \sim \chi^2(1)$, pa možemo iskoristiti funkciju izvodnicu momenata za $\chi^2(1)$*

$$\mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right] = \mathbb{E}\left[\exp\left(\frac{\sigma^2}{s^2}\left(\frac{X}{\sigma}\right)^2\right)\right] = M_{\left(\frac{X}{\sigma}\right)^2}\left(\frac{\sigma^2}{s^2}\right) = \begin{cases} \frac{1}{\sqrt{1-2\frac{\sigma^2}{s^2}}} & \frac{\sigma^2}{s^2} < \frac{1}{2} \\ \infty & \textit{inače} \end{cases}.$$

*Sada lagano kao u (a) dobijemo $\|X\|_{\psi_2} = \frac{2\sqrt{6}}{3}\sigma$.*

Iz gornjeg primjera slijedi da za nezavisne slučajne varijable $X_i \sim N(0, \sigma_i^2)$ imamo $\left\|\sum_{i=1}^n X_i\right\|_{\psi_2}^2 = \sum_{i=1}^n \|X_i\|_{\psi_2}^2$. Slična relacija vrijedi za proizvoljne nezavisne subgaussovske slučajne varijable.

**Propozicija 2.17.** *Neka su $X_1, \ldots, X_n$ nezavisne, subgaussovske slučajne varijable s očekivanjem 0. Tada je i $\sum_{i=1}^n X_i$ subgaussovska slučajna varijabla i vrijedi*

$$\left\|\sum_{i=1}^n X_i\right\|_{\psi_2}^2 \lesssim \sum_{i=1}^n \|X_i\|_{\psi_2}^2.$$

*Dokaz.* Vidi [7, Proposition 2.6.1]. □

Iz subaditivnosti norme slijedi da ako subgaussovskoj slučajnoj varijabli dodamo konstantu, rezultat će biti subgaussovska varijabla. Sljedeća lema govori nam da će prilikom centriranja slučajne varijable subgaussovska norma narasti najviše za apsolutno konstantan faktor.

**Lema 2.18** (Centriranje)**.** *Neka je $X$ subgaussovska slučajna varijabla. Tada je i $X - \mathbb{E}[X]$ subgaussovska i vrijedi*

$$\|X - \mathbb{E}[X]\|_{\psi_2} \lesssim \|X\|_{\psi_2}.$$

*Dokaz.* Vidi [7, Lemma 2.6.8]. □

Kao u Napomeni 2.15, iz svojstva (ii) iz definicije subgaussovske norme slijedi da je $\mathbb{E}[|X|] = \|X\|_{L^1} \lesssim \|X\|_{\psi_2}$. Sljedeća tvrdnja jako je korisna, jer govori da za subgaussovske varijable $X_1, X_2, \ldots$ vrijednost $\mathbb{E}[\max_{1\le i\le n}|X_i|]$ raste najviše logaritamski brzo u $n$.

**Lema 2.19.** *Neka su $X_1, X_2, \ldots$ subgaussovske slučajne varijable. Tada vrijedi*

$$\mathbb{E}\left[\sup_i \frac{|X_i|}{\sqrt{1+\log(i)}}\right] \lesssim \sup_i \|X_i\|_{\psi_2}.$$

*Posebno, za $n \ge 2$ vrijedi*

$$\mathbb{E}\left[\max_{i\le n}|X_i|\right] \lesssim \max_{i\le n}\|X_i\|_{\psi_2}\sqrt{\log(n)}.$$

*Dokaz.* Označimo $K = \max_i \|X_i\|_{\psi_2}$. Računamo

$$\begin{aligned}
\mathbb{E}\left[\sup_i \frac{|X_i|}{\sqrt{1+\log(i)}}\right] &= \int_0^\infty \mathbb{P}\left(\sup_i \frac{|X_i|}{\sqrt{1+\log(i)}} > t\right) d\lambda(t) \\
&\le t_0 + \int_{t_0}^\infty \mathbb{P}\left(\sup_i \frac{|X_i|}{\sqrt{1+\log(i)}} > t\right) d\lambda(t) \\
&\le t_0 + \int_{t_0}^\infty \left(\sum_{i=1}^\infty \mathbb{P}\left(|X_i| > t\sqrt{1+\log(i)}\right)\right) d\lambda(t) \\
&\overset{2.15}{\le} t_0 + \int_{t_0}^\infty \left(\sum_{i=1}^\infty 2\exp\left(-\frac{C^2(1+\log(i))}{K^2}t^2\right)\right) d\lambda(t) \\
&= t_0 + \frac{2K}{C}\int_{\frac{Ct_0}{K}}^\infty \left(\sum_{i=1}^\infty e^{-(1+\log(i))t^2}\right) d\lambda(t) \\
&\le t_0 + \frac{2K}{C}\left(\sum_{i=1}^\infty i^{-\left(\frac{Ct_0}{K}\right)^2}\right)\int_{\frac{Ct_0}{K}}^\infty e^{-t^2}\, d\lambda(t) \\
&\le t_0 + \frac{\sqrt{\pi}K}{C}\left(\sum_{i=1}^\infty i^{-\left(\frac{Ct_0}{K}\right)^2}\right).
\end{aligned}$$

Uzmemo li $t_0 = \sqrt{2}\frac{K}{C}$ tako da gornji red konvergira i jednak je $\sum_{i=1}^{n} i^{-2} = \frac{\pi^2}{6}$, dobijemo

$$\mathbb{E}\left[\sup_i \frac{|X_i|}{\sqrt{1+\log(i)}}\right] \leq \sqrt{2}\frac{K}{C} + \frac{\sqrt{\pi}K}{C}\frac{\pi^2}{6} = \frac{6\sqrt{2}+\pi^{5/2}}{6C} \max_i \|X_i\|_{\psi_2}.$$

Za $n \geq 2$ vrijedi $\sqrt{1+\log(n)} \leq \sqrt{1+\frac{1}{\log(2)}}\sqrt{\log(n)}$, pa je

$$\mathbb{E}\left[\max_{i\leq n}|X_i|\right] \leq \mathbb{E}\left[\max_{i\leq n} \frac{|X_i|}{\sqrt{1+\log(i)}}\right]\sqrt{1+\log(n)} \lesssim \max_{i\leq n}\|X_i\|_{\psi_2}\sqrt{\log(n)}.$$

□

Pogledajmo jednu primjenu prethodne leme.

**Primjer 2.20** (Odabir najboljeg klasifikatora)**.** *Neka je $\mathcal{Y} = \{0,1\}$, neka je $l$ funkcija 0-1 gubitka i neka su $f_1,\dots,f_N \colon \mathcal{X} \to \mathcal{Y}$ klasifikatori. Čest je problem odabir najboljeg među danim klasifikatorima $f_1,\dots,f_N$ na temelju konačnog uzorka. Pitamo se što možemo reći o riziku klasifikatora koji najbolje izgleda na uzorku veličine $n$, odnosno o klasifikatoru s najmanjim empirijskim rizikom. Još u prvom poglavlju smo na primjeru vidjeli kako minimizacija empirijskog rizika na prevelikim familijama funkcija može loše utjecati na rizik odabrane funkcije veze. Da bismo usporedili empirijski rizik i rizik najboljeg klasifikatora, treba nam ograda koja je uniformna po svim $f_i$. Primijetimo da iz Hoeffdingove leme dobijemo*

$$\mathbb{E}\left[e^{s(R_n(f_i)-R(f_i))}\right] = \prod_{i=1}^{n}\mathbb{E}\left[s\exp\left(\frac{l(f_i(X_i),Y_i)-R(f_i)}{n}\right)\right] \leq \prod_{i=1}^{n}\exp\left(\frac{s^2}{8n^2}\right) = \exp\left(\frac{s^2}{8n}\right).$$

*Kako je $\mathbb{E}[R_n(f_i)-R(f_i)] = 0$, slično kao u Napomeni 2.15 dobijemo $\|R_n(f_i)-R(f_i)\|_{\psi_2} \lesssim \sqrt{\frac{1}{8n}}$. Sada iz prethodne leme slijedi uniformna ograda*

$$\mathbb{E}\left[\max_{1\leq i\leq N}|R_n(f_i)-R(f_i)|\right] \lesssim \sqrt{\frac{\log(N)}{n}}.$$

*Gornja ograda daje jako koristan rezultat. Ako imamo dovoljno velik uzorak $n$, možemo usporediti jako velik broj klasifikatora $N$. Dok god $N$ nije "eksponencijalan" u $n$, empirijski rizik najboljeg klasifikatora bit će dovoljno dobra aproksimacija njegovog rizika.*

**Teorem 2.21** (Hoeffdingova nejednakost)**.** *Neka su $X_1,\dots,X_n$ nezavisne, subgaussovske slučajne varijable s očekivanjem 0. Tada za svaki $t \geq 0$ vrijedi*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2\exp\left(-\frac{Ct^2}{\sum_{i=1}^{n}\|X_i\|_{\psi_2}^2}\right),$$

*gdje je $C > 0$ apsolutna konstanta.*

*Dokaz.* Slijedi iz Napomene 2.15 i Propozicije 2.17

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \overset{2.15}{\leq} 2\exp\left(-\frac{C_1 t^2}{\left\|\sum_{i=1}^{n} X_i\right\|_{\psi_2}^2}\right) \overset{2.17}{\leq} 2\exp\left(-\frac{(C_1 C_2)t^2}{\sum_{i=1}^{n}\|X_i\|_{\psi_2}^2}\right). \qquad \square$$

Usporedimo li ovu verziju Hoeffdingove nejednakosti s Centralnim graničnim teoremom, za jednako distribuirane $X_i$ dobivamo

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\sqrt{n}\right) \leq 2\exp\left(-\frac{(C_1 C_2)t^2}{\|X_1\|_{\psi_2}^2}\right).$$

Vidimo da smo dobili isti rezultat kao kod klasične Hoeffdingove nejednakosti: "tipične" devijacije su najviše reda veličine $\sqrt{n}$ i ograda na rep je subgaussovska.

## 2.3 Subeksponencijalne slučajne varijable

Vidjeli smo da Bernsteinova nejednakost za velike devijacije daje eksponencijalnu ogradu na rep oblika $e^{-ct}$. Slučajne varijable s takvim repom su prirodna pojava ako promatramo subgaussovske varijable. Na primjer, ako je $X$ subgaussovska takva da je $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2}$ slučajna varijabla, onda vrijedi

$$\mathbb{P}\left(X^2 \geq t\right) = \mathbb{P}\left(|X| \geq \sqrt{t}\right) \leq 2e^{-c(\sqrt{t})^2} = 2e^{-ct},$$

pa vidimo da su repovi od $X^2$ manji od repova eksponencijalne distribucije. Takve slučajne varijable zvat ćemo subeksponencijalne.

**Definicija 2.22** (Subeksponencijalna slučajna varijabla)**.** *Kažemo da je slučajna varijabla X subeksponencijalna ako zadovoljava neko od sljedećih ekvivalentnih svojstava. Pozitivne konstante $K_i$ koje se pojavljuju u donjim svojstvima razlikuju se najviše za faktor koji je apsolutna konstanta.*

*(i) Repovi od X zadovoljavaju*

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t}{K_1}\right) \quad \text{za svaki } t \geq 0.$$

*(ii) Momenti od X zadovoljavaju*

$$\|X\|_{L^p} = \mathbb{E}[|X|^p]^{\frac{1}{p}} \leq K_2 p \quad \text{za svaki } p \geq 1.$$

*(iii) Funkcija izvodnica momenata od* $|X|$ *zadovoljava*

$$\mathbb{E}\Big[e^{s|X|}\Big] \leq e^{K_3 s} \quad \text{za svaki } s \in \left[0, \frac{1}{K_3}\right].$$

*(iv) Funkcija izvodnica momenata od* $|X|$ *zadovoljava*

$$\mathbb{E}\left[\exp\left(\frac{|X|}{K_4}\right)\right] \leq 2.$$

*Ako je* $\mathbb{E}[X] = 0$*, onda su svojstva (i)-(iv) ekvivalentna sljedećem svojstvu.*

*(v) Funkcija izvodnica momenata od* $X$ *zadovoljava*

$$\mathbb{E}\Big[e^{sX}\Big] \leq e^{K_5^2 s^2}, \quad \text{za svaki } s \text{ takav da je } |s| \leq \frac{1}{K_5}.$$

*Dokaz ekvivalencije svojstava (i)-(v) sličan je dokazu za subgaussovske i dio dokaza je dan u [7, Proposition 2.5.7].*

Napomenimo da se u literatoru (vidi [3]) subeksponencijalnim distribucijama još nazivaju funkcije distribucije $F$ za koje vrijedi

$$\lim_{x \to \infty} \frac{1 - F^{n*}(x)}{1 - F(x)} = n, \quad \text{za svaki (ili neki) } n \geq 2,$$

gdje smo sa $F^{n*}$ označili $n$-struku konvoluciju funkcije distribucije $F$. Za takve distribucije vrijedi

$$\lim_{x \to \infty} \frac{1 - F(x)}{e^{-\varepsilon x}} = \infty, \quad \text{za svaki } \varepsilon > 0,$$

što znači da imaju repove koji opadaju sporije od svake eksponencijalne razdiobe. Jasno je da takve distribucije ne zadovoljavaju svojstvo (i) Definicije 2.22 subeksponencijalnih distribucija kojima ćemo se mi baviti.

Opet možemo promatrati najmanje $K_i$ za koje vrijede svojstva iz definicije i definirati subeskponencijalnu normu.

**Definicija 2.23** (Subeksponencijalna norma)**.** *Neka je* $X$ *subeksponencijalna slučajna varijabla. Subeksponencijalnu normu definiramo kao najmanji* $K_4$ *iz svojstva (iv) iz prethodne definicije*

$$\|X\|_{\psi_1} = \inf\left\{ s > 0 \colon \mathbb{E}\left[\exp\left(\frac{|X|}{s}\right)\right] \leq 2 \right\}.$$

**Napomena 2.24.** *Svojstva (i)-(v) iz Definicije 2.22 možemo zapisati preko subeksponencijalne norme. Na primjer ako je* $\mathbb{E}[X] = 0$*, vrijedi*

$$\mathbb{E}\Big[e^{sX}\Big] \leq e^{C^2 s^2 \|X\|_{\psi_1}^2} \qquad \text{za svaki } s \text{ takav da je } |s| \leq \frac{1}{C\|X\|_{\psi_1}},$$

*gdje je* $C > 0$ *apsolutna konstanta. Također vrijedi da je* $\|X\|_{\psi_1}$*, do na apsolutno konstantan faktor, najmanji broj za kojeg vrijedi gornja nejednakost.*

Navedimo nekoliko korisnih svojstava subeksponencijalnih slučajnih varijabli.

(i) Za slučajnu varijablu $X$ vrijedi

$$\left\|X^2\right\|_{\psi_1} = \inf\left\{s > 0\colon \mathbb{E}\left[\exp\left(\frac{X^2}{s}\right)\right] \leq 2\right\} = \inf\left(\left\{s > 0\colon \mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right] \leq 2\right\}\right)^2 = \|X\|_{\psi_2}^2,$$

pa je $X$ subgaussovska ako i samo ako je $X^2$ subeksponencijalna.

(ii) Za subgaussovske slučajne varijable $X$ i $Y$ vrijedi

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

Za dokaz vidi [7, Lemma 2.7.7].

(iii) Kao u Primjeru 2.16, lako možemo pokazati da su sve ograničene slučajne varijable subeskponencijanle i vrijedi

$$\|X\|_{\psi_1} \lesssim \|X\|_\infty.$$

(iv) Za subeksponencijalnu varijablu $X$ iz (iii) slijedi $\|\mathbb{E}[X]\|_{\psi_1} \lesssim |\mathbb{E}[X]| \leq \mathbb{E}[|X|] \lesssim \|X\|_{\psi_1}$, pa za centriranu varijablu $X$ vrijedi

$$\|X - \mathbb{E}[X]\|_{\psi_1} \leq \|X\|_{\psi_1} + \|\mathbb{E}[X]\|_{\psi_1} \leq \|X\|_{\psi_1} + C\|X\|_{\psi_1} \lesssim \|X\|_{\psi_1}$$

Pokažimo sada Bernsteinovu nejednakost za subeksponencijalne varijable.

**Teorem 2.25** (Bernsteinova nejednakost)**.** *Neka su* $X_1, \ldots, X_n$ *nezavisne subeksponencijalne slučajne varijable s očekivanjem* $0$*. Tada za svaki* $t \geq 0$ *vrijedi*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2\exp\left(-C\min\left\{\frac{t^2}{\sum_{i=1}^n \|X_i\|_{\psi_1}^2}, \frac{t}{\max_{i\leq n}\|X_i\|_{\psi_1}}\right\}\right),$$

*gdje je* $C > 0$ *apsolutna konstanta.*

*Dokaz.* Koristeći Chernoffljevu ogradu i nezavisnost, za $t \geq 0$ i $s \in \left[0, \frac{1}{C \max_{i \leq n} \|X_i\|_{\psi_1}}\right]$ dobivamo

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq e^{-st}\mathbb{E}\left[e^{s \sum_{i=1}^{n} X_i}\right] = e^{-st} \prod_{i=1}^{n} \mathbb{E}\left[e^{sX_i}\right] \overset{2.24}{\leq} \exp\left(C^2 s^2 \sum_{i=1}^{n} \|X\|_{\psi_1}^2 - st\right)$$

Minimizacijom kvadratne funkcije u eksponentu po $s$ uz uvjet $s \in \left[0, \frac{1}{C \max_{i \leq n} \|X_i\|_{\psi_1}}\right]$ dobijemo da se minimum postiže u

$$s^* = \min\left\{\frac{t}{2C^2 \sum_{i=1}^{n} \|X_i\|_{\psi_1}^2}, \frac{1}{C \max_{i \leq n} \|X_i\|_{\psi_1}}\right\}.$$

Ako je $s^* = \frac{t}{2C^2 \sum_{i=1}^{n} \|X_i\|_{\psi_1}^2}$, uvrštavanjem dobijemo

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(-\frac{t^2}{4C^2 \sum_{i=1}^{n} \|X_i\|_{\psi_1}^2}\right).$$

Ako je $s^* = \frac{1}{C \max_{i \leq n} \|X_i\|_{\psi_1}}$, onda je $\sum_{i=1}^{n} \|X_i\|_{\psi_1}^2 \leq \frac{t}{2C} \max_{i \leq n} \|X_i\|_{\psi_1}$, pa vrijedi

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(\frac{\sum_{i=1}^{n} \|X_i\|_{\psi_1}^2}{\max_{i \leq n} \|X_i\|_{\psi_1}^2} - \frac{t}{C \max_{i \leq n} \|X_i\|_{\psi_1}}\right) \leq \exp\left(-\frac{t}{2C \max_{i \leq n} \|X_i\|_{\psi_1}}\right).$$

Spajanjem dobivenih ograda na rep dobijemo

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(-\min\left\{\frac{1}{4C^2}, \frac{1}{2C}\right\} \cdot \min\left\{\frac{t^2}{\sum_{i=1}^{n} \|X_i\|_{\psi_1}^2}, \frac{t}{\max_{i \leq n} \|X_i\|_{\psi_1}}\right\}\right).$$

Ponavljanjem gornjeg postupka za $(-X_i)_{i=1}^{n}$ dobijemo istu ogradu za $\mathbb{P}(-\sum_{i=1}^{n} X_i \geq t)$, pa slijedi tvrdnja teorema. □

Ako zbog jednostavnosti pretpostavimo $\|X_1\|_{\psi_1} = \cdots = \|X_n\|_{\psi_1}$, Bernsteinovu nejednakost još možemo zapisati kao

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\sqrt{n}\right) \leq 2\exp\left(-C \min\left\{\frac{t^2}{\|X_1\|_{\psi_1}^2}, \frac{\sqrt{n}t}{\|X_1\|_{\psi_1}}\right\}\right).$$

Opet vidimo da su za male devijacije $t \leq C'\sqrt{n}$ repovi sume subgaussovski

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\sqrt{n}\right) \leq 2\exp\left(-\frac{C}{\max\left\{\|X_1\|_{\psi_1}^2, C'\|X_1\|_{\psi_1}\right\}} t^2\right),$$

dok su za velike devijacije $t \geq C' \sqrt{n}$ repovi subeksponencijalni

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\sqrt{n}\right) \leq 2\exp\left(-\frac{C}{\max\left\{\frac{\|X_1\|_{\psi_1}^2}{C'}, \|X_1\|_{\psi_1}\right\}}\sqrt{n}t\right).$$

# Poglavlje 3

# Martingalna metoda

U ovom poglavlju bavimo se koncentracijskim nejednakostima za općenite funkcije nezavisnih slučajnih varijabli. Preciznije, bavit ćemo se koncentracijskim nejednakostima za slučajnu varijablu $Z := g(X_1, \dots, X_n)$, gdje su $X_1, \dots, X_n$ nezavisne slučajne varijable i $g \colon \mathbb{R}^n \to \mathbb{R}$ Borelova funkcija. Rezultati u ovom poglavlju zasnivat će se na rastavu

$$Z - \mathbb{E}[Z] = \sum_{i=1}^{n} \underbrace{(\mathbb{E}[Z \mid X_1, \dots, X_i] - \mathbb{E}[Z \mid X_1, \dots, X_{i-1}])}_{=:V_i}.$$

Intuitivno, slučajna varijabla $V_i$ daje informaciju o utjecaju varijable $X_i$ na $Z$. Gornji rastav omogućit će nam da varijabilnost slučajne varijable $Z$ ograničimo pomoću utjecaja pojedinih varijabli $X_i$ na $Z$.

Radi kraćeg zapisa, malim podebljanim simbolima označavat ćemo $n$-dimenzionalne vektore kao $\mathbf{x} = (x_1, \dots, x_n)$, a velikim slučajne vektore kao $\mathbf{X} = (X_1, \dots, X_n)$. Za $i \in \{1, \dots, n\}$ još uvodimo oznake za vektore bez $i$-te komponente

$$\mathbf{x}^{(-i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$
$$\mathbf{X}^{(-i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$$

i oznake za vektore s izmijenjenom $i$-tom komponentom

$$\mathbf{x}^{(i)} = (x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n)$$
$$\mathbf{X}^{(i)} = (X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n).$$

## 3.1 Efron-Steinova nejednakost

Efron-Steinova nejednakost daje ogradu za varijancu funkcije nezavisnih slučajnih varijabli $Z = g(X_1, \dots, X_n)$. Iz nje moći ćemo dobiti i ograde za rep, na primjer pomoću Čebiševljeve nejednakosti. U dokazu tvrdnji će nam trebati sljedeća tehnička lema, koja govori da iz uvjetnog očekivanja možemo izbaciti varijable koje su nezavisne od svega ostalog.

**Lema 3.1.** *Neka je X slučajna varijabla na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, \mathbb{P})$ *i* $\mathbb{E}[|X|] < \infty$. *Neka su* $\mathcal{G}$ *i* $\mathcal{H}$ *podalgebre od* $\mathcal{F}$ *takve da su* $\sigma(\sigma(X) \cup \mathcal{G})$ *i* $\mathcal{H}$ *nezavisne. Tada vrijedi*

$$\mathbb{E}[X \mid \sigma(\mathcal{G} \cup \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}].$$

*Dokaz.* Dovoljno je pokazati da vrijedi

$$\mathbb{E}[X \cdot \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_A]. \tag{3.1}$$

za sve $A \in \sigma(\mathcal{G} \cup \mathcal{H})$. Pokažimo prvo da jednakost vrijedi za sve skupove oblika $G \cap H$, $G \in \mathcal{G}$, $H \in \mathcal{H}$. To će vrijediti zbog nezavisnosti

$$\begin{aligned}\mathbb{E}[X \cdot \mathbb{1}_{G\cap H}] &= \mathbb{E}[(X \cdot \mathbb{1}_G) \cdot \mathbb{1}_H] = \mathbb{E}[X \cdot \mathbb{1}_G] \cdot \mathbb{E}[\mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \cdot \mathbb{1}_G \mid \mathcal{G}]] \cdot \mathbb{E}[\mathbb{1}_H] \\ &= \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_G] \cdot \mathbb{E}[\mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_G \cdot \mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_{G\cap H}].\end{aligned}$$

Neka je $\Pi = \{G \cap H \colon G \in \mathcal{G}, H \in \mathcal{H}\}$ i $\mathcal{A} = \{A \in \sigma(\mathcal{G} \cup \mathcal{H}) \colon \text{A zadovoljava (3.1)}\}$. Tada je $\Pi$ generirajući $\pi$-sustav za $\sigma(\mathcal{G} \cup \mathcal{H})$, pa je $\mathcal{D}(\Pi) = \sigma(\mathcal{G} \cup \mathcal{H})$ po Dynkinovom teoremu, gdje je $\mathcal{D}(\Pi)$ najmanja Dynkinova klasa koja sadrži $\Pi$. Također imamo

$$\begin{aligned}\Pi \subseteq \mathcal{A} \subseteq \sigma(\mathcal{G} \cup \mathcal{H}) &\implies \mathcal{D}(\Pi) \subseteq \mathcal{D}(\mathcal{A}) \subseteq \sigma(\mathcal{G} \cup \mathcal{H}) = \mathcal{D}(\Pi) \\ &\implies \mathcal{D}(\mathcal{A}) = \sigma(\mathcal{G} \cup \mathcal{H}).\end{aligned}$$

Zato je dovoljno pokazati da je $\mathcal{A}$ Dynkinova klasa, jer će tada biti $\mathcal{A} = \sigma(\mathcal{G} \cup \mathcal{H})$. Provjerimo svojstva Dynkinove klase.

(i) $\Omega = \Omega \cap \Omega \in \Pi \implies \Omega \in \mathcal{A}$.

(ii) Ako su $A, B \in \mathcal{A}$, $B \subseteq A$, tada je očito

$$\begin{aligned}\mathbb{E}[X \cdot \mathbb{1}_{A\setminus B}] &= \mathbb{E}[X \cdot (\mathbb{1}_A - \mathbb{1}_B)] = \mathbb{E}[X \cdot \mathbb{1}_A] - \mathbb{E}[X \cdot \mathbb{1}_B] \\ &\overset{(3.1)}{=} \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_A] - \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_B] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot (\mathbb{1}_A - \mathbb{1}_B)] \\ &= \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_{A\setminus B}],\end{aligned}$$

pa je $A \setminus B \in \mathcal{A}$.

(iii) Neka su $A_1 \subseteq A_2, \subseteq \cdots \in \mathcal{A}$. Tada vrijedi $A_n \uparrow \cup_{n\in\mathbb{N}} A_n$. Također, $X \cdot \mathbb{1}_{A_n}$ i $\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_{A_n}$ su dominirane sa $|X|$ za svaki $n$, pa vrijedi

$$\begin{aligned}\mathbb{E}[X \cdot \mathbb{1}_{\cup_{n\in\mathbb{N}} A_n}] &= \mathbb{E}\Big[\lim_{n\to\infty} X \cdot \mathbb{1}_{A_n}\Big] \overset{\text{LTDK}}{=} \lim_{n\to\infty} \mathbb{E}[X \cdot \mathbb{1}_{A_n}] \overset{(3.1)}{=} \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_{A_n}] \\ &\overset{\text{LTDK}}{=} \mathbb{E}[\lim_{n\to\infty} \mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_{A_n}] \overset{(3.1)}{=} \mathbb{E}[\lim_{n\to\infty} \mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{1}_{\cup_{n\in\mathbb{N}} A_n}],\end{aligned}$$

pa je $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$. $\square$

Sljedeći teorem daje ogradu na varijabilnost od $Z = g(X_1, \ldots, X_n)$ preko varijabilnosti u svakoj pojedinoj varijabli $X_i$.

**Teorem 3.2.** *Neka su $X_1, \ldots, X_n$ nezavisne slučajne varijable i neka je $g\colon \mathbb{R}^n \to \mathbb{R}$ Borelova funkcija. Uvedimo oznake $Z := g(\mathbf{X})$ i $\mathbb{E}_i[\cdot] := \mathbb{E}\big[\cdot \bigm| \mathbf{X}^{(-i)}\big]$. Ako $Z$ ima konačnu varijancu, vrijedi*

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\Big[(Z - \mathbb{E}_i[Z])^2\Big].$$

*Dokaz.* Konačnost varijance (pa i očekivanja) od $Z$ trebat će nam samo da bi sva uvjetna očekivanja bila dobro definirana. Uvedimo oznaku $V := Z - \mathbb{E}[Z]$ i definirajmo

$$V_i := \mathbb{E}[Z \mid X_1, \ldots, X_i] - \mathbb{E}[Z \mid X_1, \ldots, X_{i-1}], \qquad i = 1, \ldots, n.$$

Zbog $V_n = Z - \mathbb{E}[Z \mid X_1, \ldots, X_{n-1}]$ i $V_1 = \mathbb{E}[Z \mid X_1] - \mathbb{E}[Z]$ slijedi

$$\sum_{i=1}^{n} V_i = Z - \mathbb{E}[Z] = V.$$

Također, za $i < j$ vrijedi

$$\begin{aligned}
\mathbb{E}_i\big[V_i V_j\big] &= \mathbb{E}\big[V_i V_j \bigm| X_1, \ldots, X_i\big] \\
&= V_i \cdot \mathbb{E}\big[V_j \bigm| X_1, \ldots, X_i\big] \\
&= V_i \cdot \mathbb{E}\big[\mathbb{E}\big[Z \bigm| X_1, \ldots, X_j\big] - \mathbb{E}\big[Z \bigm| X_1, \ldots, X_{j-1}\big] \bigm| X_1, \ldots, X_i\big] \\
&= V_i \cdot (\mathbb{E}[Z \mid X_1, \ldots, X_i] - \mathbb{E}[Z \mid X_1, \ldots, X_i]) \\
&= 0.
\end{aligned}$$

Računanjem očekivanja dobivamo $\mathbb{E}\big[V_i V_j\big] = 0$ za $i < j$. Sada imamo

$$\mathrm{Var}(Z) = \mathrm{Var}(V) = \mathbb{E}\big[V^2\big] = \mathbb{E}\left[\left(\sum_{i=1}^{n} V_i\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}\big[V_i^2\big] + 2 \sum_{i<j} \underbrace{\mathbb{E}\big[V_i V_j\big]}_{0} = \sum_{i=1}^{n} \mathbb{E}\big[V_i^2\big].$$

Kako je $\sigma(\mathbb{E}_i[Z], X_1, \ldots, X_{i-1}) \subseteq \sigma(X_j\colon j \neq i)$, slijedi da su $\sigma(\mathbb{E}_i[Z], X_1, \ldots, X_{i-1})$ i $\sigma(X_i)$ nezavisne, pa iz prethodne leme slijedi

$$\mathbb{E}[\mathbb{E}_i[Z] \bigm| X_1, \ldots, X_{i-1}, X_i] \overset{3.1}{=} \mathbb{E}[\mathbb{E}_i[Z] \bigm| X_1, \ldots, X_{i-1}] = \mathbb{E}[Z \mid X_1, \ldots, X_{i-1}]. \tag{3.2}$$

Sada imamo

$$\begin{aligned} V_i^2 &= (\mathbb{E}[Z \mid X_1, \dots, X_i] - \mathbb{E}[Z \mid X_1, \dots, X_{i-1}])^2 \\ &\overset{(3.2)}{=} \Big(\mathbb{E}[Z \mid X_1, \dots, X_i] - \mathbb{E}\big[\mathbb{E}_i[Z] \,\big|\, X_1, \dots, X_i\big]\Big)^2 \\ &= \mathbb{E}\big[Z - \mathbb{E}_i[Z] \,\big|\, X_1, \dots, X_i\big]^2 \\ &\leq \mathbb{E}\Big[(Z - \mathbb{E}_i[Z])^2 \,\Big|\, X_1, \dots, X_i\Big]. \end{aligned}$$

Gdje zadnja nejednakost slijedi iz Jensenove nejednakosti za uvjetno očekivanje. Sada računanjem očekivanja dobivamo

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} \mathbb{E}\Big[V_i^2\Big] \leq \sum_{i=1}^{n} \mathbb{E}\Big[\mathbb{E}\Big[(Z - \mathbb{E}_i[Z])^2 \,\Big|\, X_1, \dots, X_i\Big]\Big] = \sum_{i=1}^{n} \mathbb{E}\Big[(Z - \mathbb{E}_i[Z])^2\Big]. \qquad \square$$

Lako pokažemo da nejednakost iz prethodnog teorema vrijedi i ako $\mathbb{E}_i[Z]$ zamijenimo s proizvoljnom $\sigma(X_j \colon j \neq i)$-izmjerivom slučajnom varijablom.

**Korolar 3.3.** *Uz oznake kao u prethodnom teoremu, neka su $Z_1, \dots, Z_n$ slučajne varijable s konačnom varijancom takve da je $Z_i$ izmjeriva u odnosu na $\sigma(X_j \colon j \neq i)$, za svaki $i = 1, \dots, n$. Tada vrijedi*

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\Big[(Z - Z_i)^2\Big].$$

*Dokaz.* Prisjetimo se da je za integrabilnu slučajnu varijablu uvjetno očekivanje ujedno i projekcija na odgovarajući potprostor. Konkretno, za svaki $i = 1, \dots, n$ vrijedi da je $Z_i'$ projekcija slučajne varijable $Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ na potprostor $L^2(\Omega, \sigma(X_j \colon j \neq i), \mathbb{P})$. Zato vrijedi

$$\mathbb{E}\Big[(Z - Z_i')^2\Big] \leq \mathbb{E}\Big[(Z - Z_i)^2\Big].$$

Sada vidimo da je nejednakost posljedica prethodnog teorema. $\square$

Iz prethodnog teorema lako slijedi i Efron-Steinova nejednakost.

**Teorem 3.4** (Efron-Stein)**.** *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable i neka su $X_1', \dots, X_n'$ nezavisne kopije od $X_1, \dots, X_n$ redom. Uz oznake kao u prethodnom teoremu i $Z_i' = g(X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n)$ vrijedi*

$$\mathrm{Var}(Z) \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\Big[(Z - Z_i')^2\Big].$$

*Dokaz.* Fiksirajmo $i \in \{1, \ldots, n\}$. Zbog nezavisnosti vrijedi da su $Z$ i $Z_i'$ uvjetno nezavisne u odnosu na $\sigma(X_j : j \neq i)$, pa vrijedi $\mathbb{E}_i[ZZ_i'] = \mathbb{E}_i[Z]\mathbb{E}_i[Z_i']$. Također, zbog nezavisnosti i jednake distribuiranosti $X_i$ i $X_i'$ vrijedi da je $\mathbb{E}_i[Z] = \mathbb{E}_i[Z_i']$ i $\mathbb{E}_i[Z^2] = \mathbb{E}_i[Z_i'^2]$. Zato je

$$\begin{aligned}\mathbb{E}_i\Big[(Z - Z_i')^2\Big] &= \mathbb{E}_i\Big[Z^2\Big] + \mathbb{E}_i\Big[Z_i'^2\Big] - 2\mathbb{E}_i[ZZ_i'] \\ &= \mathbb{E}_i\Big[Z^2\Big] + \mathbb{E}_i\Big[Z_i'^2\Big] - 2\mathbb{E}_i[Z]\mathbb{E}_i[Z_i'] \\ &= 2\mathbb{E}_i\Big[Z^2\Big] - 2\mathbb{E}_i[Z]^2 \\ &= 2\mathbb{E}_i\Big[(Z - \mathbb{E}_i[Z])^2\Big].\end{aligned}$$

Računanjem očekivanja i uvrštavanjem u prethodni teorem dobivamo

$$\mathrm{Var}(Z) \overset{3.2}{\leq} \sum_{i=1}^{n} \mathbb{E}\Big[(Z - \mathbb{E}_i[Z])^2\Big] = \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\Big[(Z - Z_i')^2\Big]. \qquad \square$$

**Napomena 3.5.** *Ako su $X_1, \ldots, X_n$ nezavisne slučajne varijable s konačnom varijancom, za $Z = \sum_{i=1}^n X_i$ postiže se jednakost u Efron-Steinovoj nejednakosti. To nam govori da su sume, u nekom smislu, minimalno koncentrirane među svim funkcijama nezavisnih slučajnih varijabli.*

## 3.2 Omeđene razlike

Vidjeli smo da varijabilnost od $Z = g(X_1, \ldots, X_n)$ možemo ograničiti preko sume varijabilnosti u svakoj pojedinoj varijabli. U ovom odjeljku bavimo se slučajem kada je varijabilnost funkcije $g$ u svakoj varijabli ograničena.

**Definicija 3.6.** *Neka je $\mathcal{X}$ skup i $g : \mathbb{R}^n \to \mathbb{R}$. Za $g$ kažemo da ima svojstvo omeđenih razlika ako postoje konstante $c_1, \ldots, c_n \geq 0$ takve da vrijedi*

$$\sup_{\substack{x_1, \ldots, x_n \in \mathbb{R} \\ x_i' \in \mathbb{R}}} \Big| g(x_1, \ldots, x_i, \ldots, x_n) - g(x_1, \ldots, x_i', \ldots, x_n) \Big| \leq c_i \quad \text{za svaki } i \in \{1, \ldots, n\}. \tag{3.3}$$

Lako vidimo da $g$ ima svojstvo omeđenih razlika ako i samo ako je ograničena, a znamo da za ograničenu slučajnu varijablu $Z$ možemo lako dobiti ogradu na varijancu i eksponencijalnu ogradu na repove. To znači da će nam svojstvo omeđenih razlika biti korisno samo ako su konstante $(c_i)_{i=1}^n$ netrivijalne. Da bismo svojstvo omeđenih razlika iskoristili na optimalan način, trebat će nam sljedeća lema.

**Lema 3.7.** *Neka su $X, Y$ nezavisni slučajni vektori s vrijednostima u $\mathbb{R}^m, \mathbb{R}^n$ redom. Neka je $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ Borelova funkcija takva da sve $x \in \mathbb{R}^m$ i $y, y' \in \mathbb{R}^n$ vrijedi*

$$|f(x, y) - f(x, y')| \leq C,$$

*za neki $C \geq 0$. Tada postoje $\sigma(X)$-izmjerive slučajne varijable $L$ i $U$ takve da vrijedi $U - L \leq C$ i $L \leq f(X,Y) \leq U$ g.s.*

*Dokaz.* U vjerojatnosnom prostoru $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_Y)$ definirajmo $l, u \colon \mathbb{R}^m \to \mathbb{R}$

$$l(x) := \operatorname*{ess\,inf}_{y \in \mathbb{R}^n} f(x,y) = \sup\{a \in \mathbb{R} \colon \mathbb{P}_Y(\{y \in \mathbb{R}^n \colon f(x,y) < a\}) = 0\} \geq \inf_{y \in R^n} f(x,y)$$
$$u(x) := \operatorname*{ess\,sup}_{y \in \mathbb{R}^n} f(x,y) = \inf\{a \in \mathbb{R} \colon \mathbb{P}_Y(\{y \in \mathbb{R}^n \colon f(x,y) > a\}) = 0\} \leq \sup_{y \in R^n} f(x,y).$$

Neka je $z \in \mathbb{R}$ proizvoljan. Zbog nezavisnosti, po Tonellijevom teoremu vrijedi da je funkcija

$$x \mapsto \int_{\mathbb{R}^n} \mathbb{1}_{\{x \in \mathbb{R}^n \colon f(x,y) > z\}} d\mathbb{P}_Y(y) = \mathbb{P}_Y(y \in \mathbb{R}^n \colon f(x,y) > z)$$

Borelova. Zato je Borelov i skup

$$\{x \in R^m \colon u(x) > c\} = \{x \in R^m \colon \mathbb{P}_Y(y \in \mathbb{R}^n \colon f(x,y) > z) > 0\}.$$

Kako je $z$ bio proizvoljan, slijedi da je $u$ Borelova. Analogno dobijemo da je i $l$ Borelova. Definirajmo $L = l(X)$ i $U = u(X)$. To su očito $\sigma(X)$-izmjerive slučajne varijable. Za svaki $x \in \mathbb{R}^m$ imamo

$$u(x) - l(x) = \operatorname*{ess\,sup}_{y \in \mathbb{R}^n} f(x,y) - \operatorname*{ess\,inf}_{y \in \mathbb{R}^n} f(x,y) \leq \sup_{y \in \mathbb{R}^n} f(x,y) - \inf_{y \in \mathbb{R}^n} f(x,y) \leq C,$$

pa je zato $U - L \leq C$. Također vrijedi

$$\mathbb{P}(f(X,Y) \leq U) = \int_{\mathbb{R}^m} \mathbb{P}_Y(y \in \mathbb{R}^n \colon f(x,y) \leq u(x)) d\mathbb{P}_X(x) = 1.$$

Analogno dobijemo i $\mathbb{P}(f(X,Y) \geq L) = 1$. □

Koristeći rezultate iz prethodnog odjeljka lako dobijemo sljedeći rezultat. Ako je $Z$ suma $n$ nezavisnih simetričnih Bernoullijevih varijabli $X_i$, lako vidimo da se postiže jednakost, pa je nejednakost oštra.

**Korolar 3.8.** *Neka funkcija $g$ ima svojstvo omeđenih razlika s konstantama $(c_i)_{i=1}^n$. Tada vrijedi*

$$\operatorname{Var}(Z) \leq \frac{1}{4} \sum_{i=1}^n c_i^2.$$

*Dokaz.* Po Lemi 3.7, za svaki $i \in \{1, \ldots, n\}$ postoje $\sigma(X_j \colon j \neq i)$-izmjerive slučajne varijable $L_i$ i $U_i$ takve da g.s. vrijedi $U_i - L_i \leq c_i$ i

$$L_i \leq g(\mathbf{X}) \leq U_i.$$

Definirajmo $Z_i = \frac{L_i+U_i}{2} \in \sigma(X_j : j \neq i)$. Iz Korolara 3.3 slijedi

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\Big[(Z - Z_i)^2\Big] \leq \sum_{i=1}^{n} \mathbb{E}\left[\left(\frac{U_i - L_i}{2}\right)^2\right] \leq \frac{1}{4}\sum_{i=1}^{n} c_i^2. \qquad \square$$

Sljedeći teorem daje eksponencijalnu ogradu za repove varijable $Z$, a ujedno je i generalizacija Hoeffdingove nejednakosti za općenite funkcije slučanih varijabli $g(X_1, \ldots, X_n)$.

**Teorem 3.9** (McDiarmid)**.** *Neka funkcija $g$ ima svojstvo omeđenih razlika s konstantama $(c_i)_{i=1}^n$. Tada za $t \geq 0$ vrijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n c_i^2}\right)$$
$$\mathbb{P}(Z - \mathbb{E}[Z] \leq -t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n c_i^2}\right).$$

*Dokaz.* Kao u dokazu Teorema 3.2, zapišimo $Z - \mathbb{E}[Z]$ kao sumu

$$Z - \mathbb{E}[Z] = \sum_{i=1}^{n} V_i, \tag{3.4}$$

gdje je

$$V_i := \mathbb{E}[Z \mid X_1, \ldots, X_i] - \mathbb{E}[Z \mid X_1, \ldots, X_{i-1}], \qquad i = 1, \ldots, n.$$

Tada je $V_i = h_i(X_1, \ldots, X_i)$ g.s., gdje je $h_i : \mathbb{R}^i \to \mathbb{R}$ Borelova funkcija dana sa

$$\begin{aligned} &h_i(x_1, \ldots, x_i) \\ &= \int_{\mathbb{R}^{n-i}} \left(\int_{\mathbb{R}} (g(x_1, \ldots, x_i, \ldots, x_n) - g(x_1, \ldots, y, \ldots, x_n)) d\mathbb{P}_{X_i}(y)\right) d\mathbb{P}_{X_{i+1}}(x_{i+1}) \ldots d\mathbb{P}_{X_n}(x_n). \end{aligned}$$

Zato je

$$\begin{aligned} &\left|h_i(x_1, \ldots, x_i) - h_i(x_1, \ldots, x_i')\right| \\ &\leq \int_{\mathbb{R}^{n-i}} \left|g(x_1, \ldots, x_i, \ldots, x_n) - g(x_1, \ldots, x_i', \ldots, x_n)\right| d\mathbb{P}_{X_{i+1}}(x_{i+1}) \ldots d\mathbb{P}_{X_n}(x_n) \\ &\leq c_i. \end{aligned}$$

Sada po Lemi 3.7 postoje slučajne varijable $L_i, U_i \in \sigma(X_1, \ldots, X_{i-1})$) takve da je $U_i - L_i \leq c_i$ i $L_i \leq V_i \leq U_i$ g.s. Kako je $\mathbb{E}[V_i \mid X_1, \ldots, X_{i-1}] = 0$, iz Leme 2.8 dobivamo da za svaki $s \in \mathbb{R}$ vrijedi

$$\mathbb{E}\Big[e^{sV_i} \;\Big|\; X_1, \ldots, X_{i-1}\Big] \leq \exp\left(\frac{s^2(U_i - L_i)^2}{8}\right) \leq \exp\left(\frac{s^2 c_i^2}{8}\right). \tag{3.5}$$

Sada za svaki $s > 0$ imamo

$$\begin{aligned}
\mathbb{E}\Big[e^{s(Z-\mathbb{E}[Z])}\Big] &\overset{(3.4)}{=} \mathbb{E}\Big[e^{s\sum_{i=1}^{n} V_i}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[e^{s\sum_{i=1}^{n} V_i} \;\Big|\; X_1, \ldots, X_{n-1}\Big]\Big] \\
&= \mathbb{E}\Big[e^{s\sum_{i=1}^{n-1} V_i}\mathbb{E}\Big[e^{sV_n} \;\Big|\; X_1, \ldots, X_{n-1}\Big]\Big] \\
&\overset{(3.5)}{\leq} \mathbb{E}\Big[e^{s\sum_{i=1}^{n-1} V_i} e^{s^2 c_n^2/8}\Big] \\
&\ldots \\
&\leq \exp\left(\frac{s^2 \sum_{i=1}^{n} c_i^2}{8}\right).
\end{aligned}$$

Preostaje još primijeniti Chernoffljevu metodu. Za $t > 0$ i $s^* = \frac{4t}{\sum_{i=1}^{n} c_i^2}$ imamo

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \overset{(2.2)}{\leq} \frac{\mathbb{E}\Big[e^{s^*(Z-\mathbb{E}[Z])}\Big]}{e^{st}} \leq \exp\left(\frac{(s^*)^2 \sum_{i=1}^{n} c_i^2}{8} - s^* t\right) = \exp\left(\frac{-2t^2}{\sum_{i=1}^{n} c_i^2}\right).$$

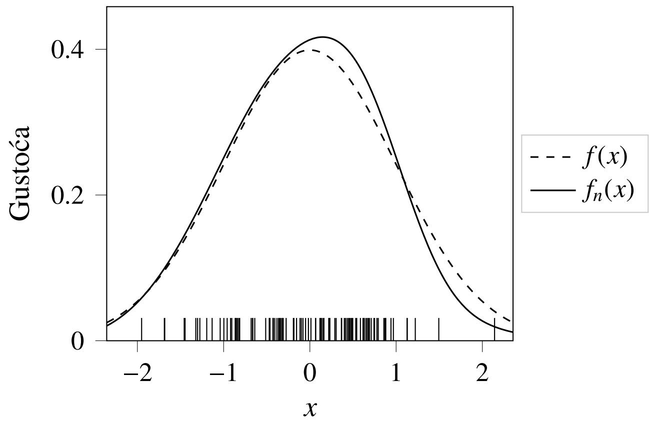
Ograda za lijevi rep slijedi ponavljanem istih argumenata za $-Z$. $\square$

Navedimo nekoliko primjena dobivenih koncentracijskih nejednakosti na funkcijama nezavisnih varijabli sa svojstvom omeđenih razlika.

**Primjer 3.10** (Procjena gustoće metodom jezgri)**.** *Neka je $X_1, \ldots, X_n$ slučajni uzorak iz distribucije s funkcijom gustoće $f$. Procjena gustoće metodom jezgri je*

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

*gdje je $K$ nenegativna izmjeriva funkcija za koju vrijedi $\int_{\mathbb{R}} K d\lambda = 1$ i $h_n > 0$ parametar zaglađivanja koji može ovisiti o veličini uzorka.*




Slika 3.1: Procjena gustoće metodom jezgri za uzorak duljine $n = 100$ iz $N(0, 1)$ koristeći normalnu jezgru $K(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$ i $h_{100} = \frac{1}{2}$.

*Kvalitetu procjene metodom jezgri možemo mjeriti $L_1$ greškom*

$$Z_n := g(X_1, \ldots, X_n) := \int_{\mathbb{R}} |f(x) - f_n(x)| d\lambda(x).$$

*Lako vidimo da vrijedi da za proizvoljan Borelov skup $A$, po točkama vrijedi*

$$\int_{\mathbb{R}} |f - f_n| d\lambda = 2 \int_{\{f > f_n\}} (f - f_n) d\lambda \geq 2 \int_{\{f > f_n\} \cap A} (f - f_n) d\lambda \geq 2 \int_A (f - f_n) d\lambda,$$

*a jednakost se postiže za $A = \{x \in \mathbb{R} \colon f(x) > f_n(x)\}$. Zato je*

$$\int_{\mathbb{R}} |f - f_n| d\lambda = 2 \sup_{A \text{ Borelov}} \left( \underbrace{\int_A f d\lambda}_{\substack{\text{Vjerojatnost} \\ \text{skupa } A}} - \underbrace{\int_A f_n d\lambda}_{\substack{\text{Procjena} \\ \text{vjerojatnosti} \\ \text{skupa } A}} \right),$$

*što pokazuje da ako je $L_1$ greška mala, procjene vjerojatnosti će biti dobre. Lako vidimo da za sve $i = 1, \dots, n$ vrijedi*

$$\begin{aligned}\left|g(\mathbf{x}) - g(\mathbf{x}^{(i)})\right| &\le \int_{\mathbb{R}} \left\| f(x) - \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - x_j}{h_n}\right) \right| - \left| f(x) - \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - \mathbf{x}_j^{(i)}}{h_n}\right) \right\| d\lambda(x) \\ &\le \frac{1}{nh_n} \int_{\mathbb{R}} \left| K\left(\frac{x - x_i}{h_n}\right) - K\left(\frac{x - x_i'}{h_n}\right) \right| d\lambda(x) \\ &\le \frac{1}{nh_n} \Bigg[ \underbrace{\int_{\mathbb{R}} K\left(\frac{x - x_i}{h_n}\right) d\lambda(x)}_{h_n} + \underbrace{\int_{\mathbb{R}} K\left(\frac{x - x_i'}{h_n}\right) d\lambda(x)}_{h_n} \Bigg] \\ &= \frac{2}{n}.\end{aligned}$$

*Sada odmah iz 3.8 dobivamo*

$$\mathrm{Var}(Z_n) \le \frac{1}{4} \sum_{i=1}^{n} \left(\frac{2}{n}\right)^2 = \frac{1}{n}, \tag{3.6}$$

*dok iz McDiarmidove nejednakosti za $t \ge 0$ slijedi*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \ge t) \le 2e^{-nt^2/2}.$$

*U [2, Chapter 5, Theorem 2] pokazano je da ako je jezgra K parna, ograničena i ima kompaktan nosač, onda vrijedi*

$$\liminf_{n\to\infty} n^{\frac{2}{5}} \mathbb{E}[Z_n] \ge C > 0,$$

*gdje je C apsolutna konstanta. Posebno, slijedi*

$$\sqrt{n}\mathbb{E}[Z_n] \xrightarrow[n\to\infty]{} \infty. \tag{3.7}$$

*Sada koristeći Čebiševljevu nejednakost, dobivamo da za svaki $\varepsilon > 0$ vrijedi*

$$\mathbb{P}\left(\left|\frac{Z_n}{\mathbb{E}[Z_n]} - 1\right| \ge \varepsilon\right) = \mathbb{P}(|Z_n - \mathbb{E}[Z_n]| \ge \varepsilon\mathbb{E}[Z_n]) \le \frac{\mathrm{Var}(Z_n)}{\varepsilon^2 \mathbb{E}[Z_n]^2} \overset{(3.6)}{\le} \frac{1}{\varepsilon^2} \cdot \frac{1}{n\mathbb{E}[Z_n]^2} \xrightarrow[n\to\infty]{(3.7)} 0.$$

*Dakle, $\frac{Z_n}{\mathbb{E}[Z_n]} \xrightarrow[n\to\infty]{} 1$ po vjerojatnosti. To svojstvo niza varijabli $(Z_n)_n$ zovemo relativna stabilnost, a interpretacija je da se $L_1$ greška asimptotski ponaša kao njeno očekivanje, pa je u određenim primjenama dovoljno analizirati očekivanja $\mathbb{E}[Z_n]$.*

**Primjer 3.11.** *Neka su $X, X_1, \dots, X_n$ nezavisne, jednako distribuirane slučajne varijable s vrijednostima u skupi $\mathcal{X}$. Neka je $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ familija podskupova od $\mathcal{X}$. Uvedimo oznaku $\mu := \mathbb{P}_X$ za distribuciju od $X$ i označimo sa $\mu_n$ empirijsku distribuciju*

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \in A\}}.$$

*Važna veličina u teoriji statistočkog učenja je*

$$Z := \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| = \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(X_i) - \mathbb{E}[\mathbb{1}_A(X)] \right|.$$

*Ako $\mathbb{P}(Z > \varepsilon) \xrightarrow[n\to\infty]{} 0$ uniformno po svim $\mu$ za svaki $\varepsilon > 0$, kažemo da je $\mathcal{A}$ uniformna Glivenko-Cantellijeva klasa (vidi Definiciju 1.9). Primijetimo da neovisno o tome kakva je familija $\mathcal{A}$, ako promijenimo vrijednost jednog $\mathcal{X}_i$, $Z$ će se promijeniti najviše za $\frac{1}{n}$, pa iz Korolara 3.8 slijedi*

$$\mathrm{Var}(Z) \leq \frac{1}{4n}.$$

*Također, iz McDiarmidove nejednakosti za $t \geq 0$ imamo eksponencijalnu ogradu na rep*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 2e^{-2nt^2}.$$

*Pokažimo sada da iz Efron-Steinove nejednakosti možemo dobiti općenitiju i bolju ogradu na varijancu. Neka je $\mathcal{F}$ familija Borelovih funkcija $f : \mathcal{X} \to \mathbb{R}$. Definirajmo* [1]

$$Z = \sup_{f \in \mathcal{F}} \left| \sum_{j=1}^{n} f(X_j) \right| = g(X_1, \dots, X_n).$$

*Primijetimo da su zbog simetrije $\left(Z - Z_i'\right)^2 \mathbb{1}_{\{Z_i' < Z\}}$ i $\left(Z_i' - Z\right)^2 \mathbb{1}_{\{Z < Z_i'\}}$ jednako distribuirane. Zato Efron-Steinovu nejednakost možemo zapisati kao*

$$\mathrm{Var}(Z) \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i')^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i')^2 \mathbb{1}_{\{Z_i' < Z\}}\right]. \tag{3.8}$$

[1] Kada radimo sa supremumom slučajnih varijabli po skupu koji može biti neprebrojiv, nećemo se baviti pitanjima izmjerivosti. Na mjestima gdje je potrebna izmjerivost od $\sup_{t \in T} X_t$ dovoljno je dodati pretpostavku o separabilnosti metričkog prostora $\left(\{X_t : t \in T\}, \|\cdot\|_{L^\infty(\mathbb{P})}\right)$, jer je onda supremum g.s. jednak supremumu po gustom prebrojivom podskupu.

*Pretpostavimo, radi jednostavnosti, da se supremum iz definicije od Z postiže u slučajnoj funkciji* $f^*$*, odnosno da je* $Z = \left|\sum_{j=1}^{n} f^*(X_j)\right|$*. Tada vrijedi*

$$\begin{aligned}
(Z - Z_i')^2 \mathbb{1}_{\{Z_i' < Z\}} &\leq \left( \left| \sum_{j=1}^{n} f^*(X_j) \right| - \left| \sum_{j \neq i} f^*(X_j) + f^*(X_i') \right| \right)^2 \mathbb{1}_{\{Z_i' < Z\}} \\
&\leq \left(f^*(X_i) - f^*(X_i')\right)^2 \mathbb{1}_{\{Z_i' < Z\}} \\
&\leq \left(f^*(X_i) - f^*(X_i')\right)^2.
\end{aligned}$$

*Zato vrijedi*

$$\begin{aligned}
\mathrm{Var}(Z) &\leq \mathbb{E}\left[ \sum_{i=1}^{n} \left(f^*(X_i) - f^*(X_i')\right)^2 \right] \\
&\leq \mathbb{E}\left[ \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \left(f(X_i) - f(X_i')\right)^2 \right] \\
&\leq \mathbb{E}\left[ \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} 2\Big(f(X_i)^2 + f(X_i')^2\Big) \right] \\
&\leq 2\mathbb{E}\left[ \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i)^2 + \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i')^2 \right] \\
&= 4\mathbb{E}\left[ \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i)^2 \right]. \qquad (3.9)
\end{aligned}$$

*Slično se pokaže da ista nejednakost vrijedi ako Z definiramo bez apsolutne vrijednosti*

$$\mathrm{Var}\left( \sup_{f \in \mathcal{F}} \sum_{j=1}^{n} f(X_j) \right) \leq 4\mathbb{E}\left[ \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i)^2 \right].$$

*Ako funkcije u* $\mathcal{F}$ *poprimaju vrijednosti u* $[-1, 1]$*, iz Korolara 3.8 smo odmah mogli zaključiti* $\mathrm{Var}(Z) \leq n$*. Kako je supremum izvan sume u* (3.9)*, nova ograda na varijancu može biti puno bolja. Dodatna prednost je da za* (3.9) *nije potrebna pretpostavka ograničenosti funkcija u* $\mathcal{F}$*.*

**Primjer 3.12** (Minimum empirijskog rizika)**.** *Prisjetimo se definicija funkcije gubitka i empirijskog rizika iz 1. poglavlja. Neka je l funkcija 0-1 gubitka. Pretpostavimo, radi jednostavnosti, da se postiže minimum empirijskog rizika*

$$\hat{R}_n = \min_{f \in \mathcal{F}} R_n(f)$$

*Neka se minimum postiže u slučajnoj funkciji $f_{\mathbf{X},\mathbf{Y}}$. Jasno je da je $\hat{R}_n$ funkcija* [2] *slučajnog uzorka*

$$\hat{R}_n = \frac{1}{n}\sum_{i=1}^{n} l(f_{\mathbf{X},\mathbf{Y}}(X_i), Y_i) =: g(\mathbf{X}, \mathbf{Y}).$$

*Promotrimo što se dogodi ako zamijenimo $X_k$ i $Y_k$ sa $X_k'$ i $Y_k'$, redom*

$$\begin{aligned} g(\mathbf{X}, \mathbf{Y}) &= \frac{1}{n}\sum_{i=1}^{n} l(f_{\mathbf{X},\mathbf{Y}}(X_i), Y_i) \le \frac{1}{n}\sum_{i=1}^{n} l(f_{\mathbf{X}^{(i)},\mathbf{Y}^{(i)}}(X_i), Y_i) \\ &= g\Big(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)}\Big) - \frac{1}{n} l(f_{\mathbf{X}^{(i)},\mathbf{Y}^{(i)}}(X_i'), Y_i') + \frac{1}{n} l(f_{\mathbf{X}^{(i)},\mathbf{Y}^{(i)}}(X_i), Y_i) \\ &\le g\Big(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)}\Big) + \frac{1}{n}. \end{aligned}$$

*Analogno imamo i*

$$g\Big(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)}\Big) \le g(\mathbf{X}, \mathbf{Y}) + \frac{1}{n}.$$

*Efron-Steinova nejednakost i Korolar 3.8 vrijede i za slučajne vektore, pa Korolar 3.8 možemo primijeniti na $\hat{R}_n$, koja je funkcija slučajnih vektora $(X_1, Y_1), \ldots, (X_n, Y_n)$ i dobivamo*

$$\mathrm{Var}\Big(\hat{R}_n\Big) \le \frac{1}{4n}. \tag{3.10}$$

*Slično, iz McDiarmidove nejednakosti za $t \ge 0$ imamo eksponencijalnu ogradu na rep*

$$\mathbb{P}\Big(\Big|\hat{R}_n - \mathbb{E}\Big[\hat{R}_n\Big]\Big| \ge t\Big) \le 2e^{-2nt^2}.$$

*Pokažimo sada kako iz Efron-Steinove nejednakosti možemo dobiti bolju ogradu na varijancu. Neka je $Z = n\hat{R}_n$ i neka su $Z_i'$ definirane kao u Teoremu 3.4, to jest*

$$Z_i = \min_{f\in\mathcal{F}}\left(\sum_{j\ne i} l(f(X_j), Y_j) + l(f(X_i'), Y_i')\right),$$

*gdje su $(X_1', Y_1'), \ldots, (X_n', Y_n')$ nezavisne kopije od $(X_1, Y_1), \ldots, (X_n, Y_n)$. Kao u Primjeru 3.11 dobijemo*

$$\mathrm{Var}(Z) \le \sum_{i=1}^{n} \mathbb{E}\Big[(Z - Z_i')^2 \mathbb{1}_{\{Z_i' > Z\}}\Big].$$

[2] Da bismo zadovoljili tehničke pretpostavke izmjerivosti koje će nam trebati, dovoljno je pretpostaviti da je $\mathcal{F}$ konačna.

*Označimo sa $f_n^*$ funkciju koja minimizira empirijski rizik, dakle vrijedi* $Z = \sum_{j=1}^n l(f_n^*(X_j), Y_j)$. *Primijetimo da vrijedi*

$$\begin{aligned}\{Z_i' > Z\} &\subseteq \left\{\sum_{j\neq i} l(f_n^*(X_j), Y_j) + l(f_n^*(X_i'), Y_i') > \sum_{j=1}^n l(f_n^*(X_j), Y_j)\right\} \\ &= \{l(f_n^*(X_i'), Y_i') > l(f_n^*(X_i), Y_i)\} \\ &\subseteq \{l(f_n^*(X_i), Y_i) = 0\}.\end{aligned}$$

*Sada slično kao u Primjeru 3.11 imamo*

$$(Z - Z_i')^2 \mathbb{1}_{\{Z_i' > Z\}} \leq \left(l(f_n^*(X_i), Y_i) - l(f_n^*(X_i'), Y_i')\right)^2 \mathbb{1}_{\{Z_i' > Z\}} \leq l(f_n^*(X_i'), Y_i') \mathbb{1}_{\{l(f_n^*(X_i), Y_i) = 0\}}.$$

*Zato vrijedi*

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}\left[(Z - Z_i')^2 \mathbb{1}_{\{Z_i' > Z\}}\right] &\leq \mathbb{E}\left[\sum_{i=1}^n l(f_n^*(X_i'), Y_i') \mathbb{1}_{\{l(f_n^*(X_i), Y_i) = 0\}}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^n l(f_n^*(X_i'), Y_i') \mathbb{1}_{\{l(f_n^*(X_i), Y_i) = 0\}} \,\middle|\, \mathbf{X}, \mathbf{Y}\right]\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n \mathbb{1}_{\{l(f_n^*(X_i), Y_i) = 0\}} \underbrace{\mathbb{E}\left[l(f_n^*(X_i'), Y_i') \,\middle|\, \mathbf{X}, \mathbf{Y}\right]}_{= \int_{\mathcal{X}\times\mathcal{Y}} l(f_n^*(x), y) d\mathbb{P}_{(X_i', Y_i')}(x, y) = R(f_n^*)}\right] \\ &\leq n\mathbb{E}[R(f_n^*)].\end{aligned}$$

*Dakle, iz Efron-Steinove nejednakosti slijedi*

$$\mathrm{Var}\left(\hat{R}_n\right) = \frac{1}{n^2} \mathrm{Var}(Z) \leq \frac{\mathbb{E}[R(f_n^*)]}{n}. \tag{3.11}$$

*Ova ograda na varijancu može dati znatno poboljšanje u odnosu na* (3.10) *kad god je* $\mathbb{E}[R(f_n^*)] \ll \frac{1}{4}$, *a to je često slučaj kod binarne klasifikacije. Na primjer, imamo*

$$R(f_n^*) = \hat{R}_n - (R_n(f_n^*) - R(f_n^*)) \leq \hat{R}_n + \sup_{f\in\mathcal{F}} |R_n(f) - R(f)|.$$

*Kasnije ćemo u dokazu Teorema 5.24 pokazati da vrijedi*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |R_n(f) - R(f)|\right] \lesssim \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}, \tag{3.12}$$

*kad god $\mathcal{F}$ ima konačnu VC dimenziju. Zato će vrijediti ograda*

$$\mathrm{Var}\left(\hat{R}_n\right) \overset{(3.11)}{\leq} \frac{\mathbb{E}\left[\hat{R}_n\right]}{n} + \frac{\mathbb{E}\left[\sup_{f\in\mathcal{F}} |R_n(f) - R(f)|\right]}{n} \overset{(3.12)}{\leq} \frac{\mathbb{E}\left[\hat{R}_n\right]}{n} + \frac{C\sqrt{\mathrm{vc}(\mathcal{F})}}{n^{3/2}}$$

*Kada je* $Y = \{0, 1\}$, *minimum empirijskog rizika na desnoj strani često je manji od* $\frac{1}{4}$, *a drugi član je reda veličine* $n^{-3/2}$, *pa je ograda na varijancu obično bolja od* (3.10).

## 3.3 Samoomeđujuće funkcije

Među funkcijama sa svojstvom omeđenih razlika posebno se izdvajaju samoomeđujuće funkcije. Vidjeli smo da je kod funkcija sa svojstvom omeđenih razlika varijabilnost u svakoj varijabli omeđena konstantom. Samoomeđujuće funkcije imaju dodatno svojstvo da je zbroj varijabilnosti u svim varijablama omeđen odozgo vrijednošću same funkcije.

**Definicija 3.13.** *Kažemo da izmjeriva funkcija $g\colon \mathcal{X}^n \to \mathbb{R}$ ima samoomeđujuće svojstvo ako postoje izmjerive funkcije $g_i\colon \mathcal{X}^{n-1} \to \mathbb{R}$, $i = 1, \ldots, n$ takve da za svake $x_1, \ldots, x_n \in \mathcal{X}$ vrijedi*

$$0 \le g(\mathbf{x}) - g_i\Big(\mathbf{x}^{(-i)}\Big) \le 1 \tag{3.13}$$

*i*

$$\sum_{i=1}^{n}\Big(g(\mathbf{x}) - g_i\Big(\mathbf{x}^{(-i)}\Big)\Big) \le g(\mathbf{x}). \tag{3.14}$$

Jasno je da svaka funkcija sa samoomeđujućim svojstvom ima i svojstvo omeđenih razlika s konstantama $c_i = 1$. Sljedeći korolar daje ogradu za varijancu samoomeđujućih funkcija nezavisnih slučajnih varijabli koja je često bolja od ograde iz Korolara 3.8.

**Korolar 3.14.** *Neka su $X_1, \ldots, X_n$ nezavisne slučajne varijable i neka Borelova funkcija $g\colon \mathbb{R}^n \to \mathbb{R}$ ima samoomeđujuće svojstvo. Tada za $Z = g(X_1, \ldots, X_n)$ vrijedi*

$$\mathrm{Var}(Z) \le \mathbb{E}[Z].$$

*Dokaz.* Označimo $Z_i = g_i\Big(\mathbf{X}^{(-i)}\Big)$. Tada iz definicije samoomeđujućeg svojstva imamo

$$\sum_{i=1}^{n}(Z - Z_i)^2 \le Z,$$

pa iz Korolara 3.3 slijedi tvrdnja. □

Navedimo nekoliko tipičnih primjera funkcija sa samoomeđujućim svojstvom koje se pojavljuju u teoriji statističkog učenja.

**Primjer 3.15** (Empirijski procesi). *Neka je $\mathcal{F}$ familija izmjerivih funkcija $f\colon \mathcal{X} \to [0, 1]$. U Primjeru 3.11 pokazali smo ograde za varijancu od $Z = g(\mathbf{X})$, gdje je*

$$g(\mathbf{x}) = \sup_{f\in\mathcal{F}} \sum_{j=1}^{n} f(x_j).$$

*Procese oblika* $\left(\sum_{j=1}^n f(X_j)\right)_{f\in\mathcal{F}}$ *još zovemo empirijski procesi. Neka je*

$$g_i\left(\mathbf{x}^{(-i)}\right) = \sup_{f\in\mathcal{F}} \sum_{j\neq i} f(x_j).$$

*Kako za svaku* $f \in \mathcal{F}$ *vrijedi*

$$\sum_{j\neq i} f(X_j) \leq \sum_{j=1}^n f(X_j) \leq \sum_{j\neq i} f(X_j) + 1,$$

*računanjem supremuma po svim* $f \in \mathcal{F}$ *dobijemo* (3.13). *Također vrijedi*

$$\sum_{j=1}^n f(x_j) = \frac{1}{n-1} \sum_{i=1}^n \sum_{j\neq i} f(x_j),$$

*iz čega računanjem supremuma po svim* $f \in \mathcal{F}$ *dobijemo*

$$g(\mathbf{x}) = \frac{1}{n-1} \sup_{f\in\mathcal{F}} \sum_{i=1}^n \sum_{j\neq i} f(x_j) \leq \frac{1}{n-1} \sum_{i=1}^n \sup_{f\in\mathcal{F}} \sum_{j\neq i} f(x_j) = \frac{1}{n-1} \sum_{i=1}^n g\left(\mathbf{x}^{(-i)}\right),$$

*što je ekvivalentno* (3.14). *Dakle, g ima samoomeđujuće svojstvo, pa vrijedi* $\operatorname{Var}(Z) \leq \mathbb{E}[Z]$ *što može biti bolje od ograde* $\operatorname{Var}(Z) \leq \frac{n}{4}$ *koju dobijemo iz Korolara 3.8.*

**Primjer 3.16** (Rademacherova usrednjenja)**.** *Neka su* $X_1, \ldots, X_n, \sigma_1, \ldots, \sigma_n$ *nezavisne slučajne varijable takve da* $\sigma_1, \ldots, \sigma_n$ *imaju Rademacherovu distribuciju. Neka je* $\mathcal{F}$ *familija izmjerivih funkcija* $f\colon \mathcal{X} \to [-1, 1]$. *Uvjetno Rademacherovo usrednjenje definiramo kao*

$$Z = \underbrace{\mathbb{E}\left[\sup_{f\in\mathcal{F}} \sum_{j=1}^n \sigma_j f(X_j) \,\middle|\, \mathbf{X}\right]}_{=:g(\mathbf{X})}.$$

*Za razliku od Rademacherove kompleksnosti (vidi Definiciju 1.8), Z je funkcija slučajnih varijabli* $X_1, \ldots, X_n$ *što znači da ovisi o uzorku. Pokažimo da g ima samoomeđujuće svojstvo. Definirajmo*

$$Z_i = \mathbb{E}\left[\sup_{f\in\mathcal{F}} \sum_{j\neq i} \sigma_j f(X_j) \,\middle|\, \mathbf{X}\right] \overset{3.1}{=} \underbrace{\mathbb{E}\left[\sup_{f\in\mathcal{F}} \sum_{j\neq i} \sigma_j f(X_j) \,\middle|\, \mathbf{X}^{(-i)}\right]}_{=:g_i\left(\mathbf{X}^{(-i)}\right)}.$$

*Neka je* $\boldsymbol{\alpha}^{(-i)} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) \in \{-1,1\}^{n-1}$ *proizvoljan. Očito vrijedi*

$$\sum_{j \neq i} \alpha_j f(x_j) = \frac{1}{2} \sum_{\alpha_i \in \{-1,1\}} \sum_{j=1}^{n} \alpha_j f(x_j).$$

*Izračunamo li u gornjoj jednakosti supremum po svim* $f \in \mathcal{F}$ *i zatim aritmetičku sredinu po svim* $\boldsymbol{\alpha}^{(-i)} \in \{-1,1\}^{n-1}$*, dobijemo*

$$\begin{aligned} g_i\Big(\mathbf{x}^{(-i)}\Big) &= \frac{1}{2^{n-1}} \sum_{\boldsymbol{\alpha}^{(-i)} \in \{-1,1\}^{n-1}} \sup_{f \in \mathcal{F}} \sum_{j \neq i} \alpha_j f(x_j) = \frac{1}{2^n} \sum_{\boldsymbol{\alpha}^{(-i)} \in \{-1,1\}^{n-1}} \sup_{f \in \mathcal{F}} \sum_{\alpha_i \in \{-1,1\}} \sum_{j=1}^{n} \alpha_j f(x_j) \\ &\leq \frac{1}{2^n} \sum_{\boldsymbol{\alpha} \in \{-1,1\}^n} \sup_{f \in \mathcal{F}} \sum_{j=1}^{n} \alpha_j f(x_j) = g(\mathbf{x}). \end{aligned}$$

*Također, za svaki* $\alpha \in \{-1,1\}^n$ *vrijedi*

$$\sum_{j=1}^{n} \alpha_j f(x_j) \leq \sum_{j \neq i} \alpha_j f(x_j) + 1,$$

*iz čega računanjem supremuma po svim* $f \in \mathcal{F}$ *i zatim aritmetičke sredine po svim* $\boldsymbol{\alpha} \in \{-1,1\}^n$*, dobijemo*

$$g(\mathbf{x}) = \frac{1}{2^n} \sum_{\boldsymbol{\alpha} \in \{-1,1\}^n} \sup_{f \in \mathcal{F}} \sum_{j=1}^{n} \alpha_j f(x_j) \leq \frac{1}{2^n} \sum_{\boldsymbol{\alpha} \in \{-1,1\}^n} \sup_{f \in \mathcal{F}} \sum_{j \neq i} \alpha_j f(x_j) + 1 = g_i\Big(\mathbf{x}^{(-i)}\Big) + 1.$$

*Slično kao u prethodnom primjeru, imamo*

$$\sum_{j=1}^{n} \alpha_j f(x_j) = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{j \neq i} \alpha_j f(x_j),$$

*pa opet računanjem supremuma po svim* $f \in \mathcal{F}$ *i zatim aritmetičke sredine po svim* $\boldsymbol{\alpha} \in \{-1,1\}^n$ *dobijemo*

$$\begin{aligned} g(\mathbf{x}) &= \frac{1}{n-1} \cdot \frac{1}{2^n} \sum_{\boldsymbol{\alpha} \in \{-1,1\}^n} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \sum_{j \neq i} \alpha_j f(x_j) \leq \frac{1}{n-1} \cdot \frac{1}{2^n} \sum_{\boldsymbol{\alpha} \in \{-1,1\}^n} \sum_{i=1}^{n} \sup_{f \in \mathcal{F}} \sum_{j \neq i} \alpha_j f(x_j) \\ &= \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{1}{2^n} \sum_{\boldsymbol{\alpha} \in \{-1,1\}^n} \sup_{f \in \mathcal{F}} \sum_{j \neq i} \alpha_j f(x_j) \right) = \frac{1}{n-1} \sum_{i=1}^{n} g\Big(\mathbf{x}^{(-i)}\Big). \end{aligned}$$

*Dakle, g ima samoomeđujuće svojstvo pa iz Korolara 3.8 i 3.14 slijedi* $\mathrm{Var}(Z) \leq \frac{n}{4}$ *i* $\mathrm{Var}(Z) \leq \mathbb{E}[Z]$. *Posebno, ako su* $X_1, \ldots, X_n$ *jednako distribuirane i ako je* $\mathcal{F} \subseteq \{0,1\}^{\mathcal{X}}$ *familija klasifikatora konačne VC dimenzije* $\mathrm{vc}(\mathcal{F}) \geq 1$, *iz Teorema 5.21 će slijediti*

$$\mathrm{Var}(Z) \leq \mathbb{E}[Z] = \mathbb{E}\left[\sup_{f \in \mathcal{F}} \sum_{j=1}^{n} \sigma_j f(X_j)\right] \leq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \left|\sum_{j=1}^{n} \sigma_j f(X_j)\right|\right] \overset{5.21}{\lesssim} \sqrt{\mathrm{vc}(\mathcal{F})n},$$

*što je za veće uzorke puno bolja ograda od* $\mathrm{Var}(Z) \leq \frac{n}{4}$.

Jedna klasa samoomeđujućih funkcija koja se pojavljuje u brojnim primjenama su takozvane konfigurancijske funkcije, a tipičan je primjer VC dimenzija u odnosu na uzorak.

**Definicija 3.17** (Konfiguracijska funkcija)**.** *Neka je* $\mathcal{X}$ *skup. Svojstvo P definiramo preko niza skupiva* $(P_n)_{n=1}^{\infty}$, $P_n \subseteq \mathcal{X}^n$ *i kažemo da niz* $(x_1, \ldots, x_n) \in \mathcal{X}^n$ *zadovoljava svojstvo P ako je* $(x_1, \ldots, x_n) \in P_n$. *Dodatno pretpostavljamo da ako neki niz ima svojstvo P, onda ga ima i svaki njegov podniz. Konfiguracijska funkcija* $g_n \colon \mathcal{X}^n \to \mathbb{N}_0$ *svojstva P preslikava nizove* $(x_1, \ldots, x_n)$ *u duljinu najduljeg podniza koji ima svojstvo P.*

**Korolar 3.18.** *Neka su* $X_1, \ldots, X_n$ *nezavisne slučajne varijable i neka je* $g_n$ *konfiguracijska funkcija nekog svojstva P. Tada za* $Z = g_n(\mathbf{X})$ *vrijedi*

$$\mathrm{Var}(Z) \leq \mathbb{E}[Z].$$

*Dokaz.* Vidi [4, Corollary 3]. □

**Primjer 3.19** (VC dimenzija u odnosu na uzorak)**.** *Neka je* $\mathcal{X}$ *neki skup. Za* $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$ *i* $\mathbf{x} \in \mathcal{X}^n$ *definiramo trag od* $\mathcal{A}$ *na* $\mathbf{x}$ *kao*

$$\mathrm{tr}(x_1, \ldots, x_n) = \{A \cap \{x_1, \ldots, x_n\} \colon A \in \mathcal{A}\}.$$

*Definiramo još i* $T(\mathbf{x}) = |\mathrm{tr}(\mathbf{x})|$. *Kažemo da se* $\mathbf{x}$ *može rastaviti sa* $\mathcal{A}$ *ako je* $T(\mathbf{x}) = 2^n$. *VC dimenziju od A u odnosu na* $\mathbf{x}$, *oznaka* $D(\mathbf{x})$, *definiramo kao duljinu najvećeg podniza od* $(x_1, \ldots, x_n)$ *koji se može rastaviti sa* $\mathcal{A}$. *Primijetimo da su uvedeni pojam rastavljanja i VC dimenzije "uzoračke" verzije pojmova iz 1. poglavlja. Iz definicije je jasno da je* $D(\mathbf{x})$ *konfiguracijska funkcija za svojstvo rastavljanja. Zato za nezavisne slučajne varijable* $X_1, \ldots, X_n$ *s vrijednostima u* $\mathcal{X}$, *slučajna VC dimenzija* $D(\mathbf{X})$ *zadovoljava*

$$\mathrm{Var}(D(\mathbf{X})) \leq \mathbb{E}[D(\mathbf{X})].$$

## 3.4 Koncentracija produktne mjere

U ovom odjeljku iznosimo nekoliko rezultata koji govore o koncentraciji produktne mjere oko skupova. Rezultati iz ovog odjeljka spadaju pod izoperimetarski pristup koncentracijskim nejednakostima koji je razvio Michel Talagrand, a neke rezultate možemo pokazati i pomoću prethodnih rezultata iz ovog poglavlja. Koristit ćemo Hammingovu udaljenost.

**Definicija 3.20** (Hammingova i konveksna udaljenost)**.** *Neka je $\mathcal{X}$ skup. Hammingovu udaljenost $\mathbf{x}, \mathbf{y} \in \mathcal{X}^n$ definiramo kao broj koordinata na kojima se razlikuju*

$$d(\mathbf{x}, \mathbf{y}) = |\{i \in \{1, \ldots, n\} \colon x_i \neq y_i\}| = \sum_{i\colon x_i \neq y_i} 1.$$

*Za vektor nenegativnih težina $\boldsymbol{\alpha} \in [0, \infty\rangle^n$ definiramo težinsku Hammingovu udaljenost kao*

$$d_\alpha(\mathbf{x}, \mathbf{y}) = \sum_{i\colon x_i \neq y_i} \alpha_i.$$

*Definiramo i Hammingovu udaljenost $\mathbf{x} \in \mathcal{X}^n$ od skupa $A \subseteq \mathcal{X}^n$ kao*

$$d(\mathbf{x}, A) = \min_{\mathbf{y} \in A} d(\mathbf{x}, \mathbf{y})$$
$$d_\alpha(\mathbf{x}, A) = \min_{\mathbf{y} \in A} d_\alpha(\mathbf{x}, \mathbf{y}).$$

*Konveksnu udaljenost definiramo kao maksimalnu težinsku Hammingovu udaljenost*

$$d_T(\mathbf{x}, A) = \sup_{\alpha \in [0,\infty\rangle^n \colon \|\alpha\|_2 = 1} d_\alpha(\mathbf{x}, A).$$

**Teorem 3.21.** *Neka su $X_1, \ldots, X_n$ nezavisne slučajne varijable s vrijednostima u $\mathcal{X}$. Za svaki izmjerivi skup $A \subseteq \mathcal{X}^n$ i svaki $t > 0$ vrijedi*

$$\mathbb{P}\left( d(\mathbf{X}, A) \geq t + \sqrt{\frac{n}{2} \log\left( \frac{1}{\mathbb{P}_{\mathbf{X}}(A)} \right)} \right) \leq e^{-2t^2/n}.$$

*Dokaz.* Primijetimo da ako promijenimo jednu koordinatu od $\mathbf{x}$, vrijednost $d(\mathbf{x}, A)$ se može promijeniti najviše za 1. Zato iz Teorema 3.9 slijedi

$$\mathbb{P}(\mathbb{E}[d(\mathbf{X}, A))] - d(\mathbf{X}, A) \geq t) \leq e^{-2t^2/n}.$$

Uvrstimo li $t = \mathbb{E}[d(\mathbf{X}, A))]$, imamo

$$\mathbb{P}_{\mathbf{X}}(A) = \mathbb{P}(d(\mathbf{X}, A) \leq 0) = \mathbb{P}(\mathbb{E}[d(\mathbf{X}, A))] - d(\mathbf{X}, A) \geq \mathbb{E}[d(\mathbf{X}, A))]) \leq e^{-2\mathbb{E}[d(\mathbf{X},A))]^2/n}.$$

Iz čega slijedi

$$\mathbb{E}[d(\mathbf{X},A))] \leq \sqrt{\frac{n}{2}\log\left(\frac{1}{\mathbb{P}_{\mathbf{X}}(A)}\right)},$$

pa ponovnim korištenjem Teorema 3.9 dobivamo

$$\mathbb{P}\left(d(\mathbf{X},A) \geq t + \sqrt{\frac{n}{2}\log\left(\frac{1}{\mathbb{P}_{\mathbf{X}}(A)}\right)}\right) \leq \mathbb{P}(d(\mathbf{X},A) \geq t + \mathbb{E}[d(\mathbf{X},A))]) \leq e^{-2t^2/n}. \qquad \square$$

Uzmemo li u prethodnom teoremu skup $A$ takav da je $\mathbb{P}(A) = e^{-50}$ i $t = 5\sqrt{n}$, dobivamo

$$\mathbb{P}\left(d(\mathbf{X},A) \geq 10\sqrt{n}\right) \leq e^{-50}.$$

Gornja nejednakost za veliki $n$ daje iznenađujući rezultat, jer nam govori da su produktne mjere u visokodimenzionalnim prostorima koncentrirane oko skupova koji imaju malu vjerojatnost.

**Primjer 3.22** (Omeđene razlike i koncentracija oko medijana)**.** *Neka su $X_1, \ldots, X_n$ nezavisne slučajne varijable s vrijednostima u $\mathcal{X}$. Pretpostavimo da izmjeriva funkcija $g\colon \mathcal{X}^n \to \mathbb{R}$ ima svojstvo omeđenih razlika s konstantama $c_i = 1$, radi jednostavnosti. Neka je $A = \{\mathbf{x} \in \mathcal{X}^n : g(\mathbf{x}) \leq \mathbb{M}[Z]\}$, gdje smo sa $\mathbb{M}[Z]$ označili medijan slučajne varijable $Z = g(\mathbf{X})$. Po definiciji medijana je $\mathbb{P}_{\mathbf{X}}(A) \geq \frac{1}{2}$, pa primjenom postupka iz gornjeg dokaza na $Z - \mathbb{M}[Z]$ umjesto $d(\mathbf{X},A)$ dobivamo*

$$\mathbb{P}\left(Z - \mathbb{M}[Z] \geq t + \sqrt{\frac{n}{2}\log(2)}\right) \leq e^{-2t^2/n}.$$

*Nejednakost ima sličan oblik kao McDiarmidova nejednakost, ali s medijanom umjesto očekivanja. Koncentracija varijable oko medijana nije iznenađujuća, jer znamo da udaljenost između očekivanja i medijana nije veća od standardne devijacije*

$$|\mathbb{E}[Z] - \mathbb{M}[Z]| \leq \mathbb{E}[|Z - \mathbb{M}[Z]|] = \min_{z\in\mathbb{R}} \mathbb{E}[|Z - z|] \leq \mathbb{E}[|Z - \mathbb{E}[Z]|] \overset{\text{Jensen}}{\leq} \sqrt{\mathbb{E}\left[(Z - \mathbb{E}[Z])^2\right]}.$$

*Također, ako je $Z$ koncentrirana oko medijana, medijan će biti blizu očekivanja*

$$|\mathbb{E}[Z] - \mathbb{M}[Z]| \leq \mathbb{E}[|Z - \mathbb{M}[Z]|] = \int_0^\infty \mathbb{P}(|Z - \mathbb{M}[Z]| \geq t)dt.$$

Ponovimo li argumente iz dokaza Teorema 3.21 za $d_\alpha$, $\|\alpha\|_2 = 1$, dobijemo

$$\mathbb{P}\left(d_\alpha(\mathbf{X},A) \geq t + \sqrt{-\frac{\log(\mathbb{P}_{\mathbf{X}}(A))}{2}}\right) \leq e^{-2t^2}.$$

Označimo $u = \sqrt{-\frac{\log(\mathbb{P}_{\mathbf{X}}(A))}{2}}$. Za $t \geq u$ vrijedi

$$\mathbb{P}(d_\alpha(\mathbf{X}, A) \geq t) \leq e^{-2(t-u)^2}.$$

Primijetimo da za $t \leq 2u = \sqrt{-2\log(\mathbb{P}_{\mathbf{X}}(A))}$ vrijedi $\mathbb{P}_{\mathbf{X}}(A) \leq e^{-t^2/2}$, dok za $t \geq 2u$ vrijedi $\mathbb{P}(d_\alpha(\mathbf{X}, A) \geq t) \leq e^{-2(t-u)^2} \leq e^{-t^2/2}$. Dakle, imamo

$$\min\left\{\mathbb{P}_{\mathbf{X}}(A), \sup_{\alpha \in [0,\infty\rangle:\ \|\alpha\|_2 = 1} \mathbb{P}(d_\alpha(\mathbf{X}, A) \geq t)\right\} \leq e^{-t^2/2}.$$

Sljedeći Teorem govori da gornja nejednakost vrijedi čak i kada supremum stavimo unutar vjerojatnosti.

**Teorem 3.23** (Nejednakost konveksne udaljenosti)**.** *Neka su $X_1, \ldots, X_n$ nezavisne slučajne varijable s vrijednostima u $\mathcal{X}$ i neka je $A \subseteq \mathcal{X}^n$ izmjeriv takav da je $\mathbb{P}_{\mathbf{X}}(A) \geq \frac{1}{2}$. Za svaki $t > 0$ vrijedi*

$$\min\{\mathbb{P}_{\mathbf{X}}(A), \mathbb{P}(d_T(\mathbf{X}, A) \geq t)\} \leq e^{-t^2/4}.$$

*Dokaz.* Vidi [4, Theorem 22]. □

# Poglavlje 4

# Metoda entropije

U ovom poglavlju također ćemo se baviti funkcijama nezavisnih slučajnih varijabli $Z = g(X_1, \dots, X_n)$. Koristeći rezultate iz teorije informacija, pokazat ćemo nejednakost sličnu Efron-Steinovoj nejendakosti. Zatim ćemo metodom entropije pokazati kako se mogu dobiti eksponencijalne koncentracijske nejednakosti za $Z$. Nastavit ćemo s oznakama iz prethodnog poglavlja.

**Definicija 4.1** (Entropija)**.** *Neka je $X$ diskretna slučajna varijabla s vrijednostima u prebrojivom skupu $\mathcal{X}$. Označimo njenu funkciju gustoće sa $f_X(x) := \mathbb{P}(X = x)$. Entropija od $X$ jednaka je*

$$H(X) := \mathbb{E}[-\log(f_X(X))] = -\sum_{x\in\mathcal{X}} f_X(x)\log(f_X(x)),$$

*gdje računamo $0 \cdot \log(0) = 0$, jer je $\lim_{x\to 0+} x\log(x) = 0$. Entropiju diskretnog slučajnog vektora $(X, Y)$ označavamo sa $H(X, Y)$.*

**Definicija 4.2** (Relativna entropija)**.** *Neka su $P$ i $Q$ vjerojatnosti na prebrojivom skupu $\mathcal{X}$. Relativnu entropiju između $P$ i $Q$ definiramo* [1] *kao*

$$D(P \parallel Q) = \sum_{x\in\mathcal{X}} f_P(x)\log\left(\frac{f_P(x)}{f_Q(x)}\right).$$

Znamo da je $\log(x) \le x - 1$ i jednakost se postiže samo u $x = 1$. Zato vrijedi

$$D(P \parallel Q) = -\sum_{x\in\mathcal{X}} f_P(x)\log\left(\frac{f_Q(x)}{f_P(x)}\right) \ge -\sum_{x\in\mathcal{X}} f_P(x)\left(\frac{f_Q(x)}{f_P(x)} - 1\right) = 0,$$

[1] Definicija ima smisla samo ako je $f_P(x) = 0$ kad god je $f_Q(x) = 0$, što znači da vjerojatnost $P$ mora biti apsolutno neprekidna u odnosu na $Q$.

pa je relativna entropija nenegativna i jednaka nuli ako i samo ako je $P = Q$. Posebno, ako su $X$ i $Y$ slučajne varijable s vrijednostima na konačnom skupu $\mathcal{X}$ i distribucija od $Y$ je uniformna na $\mathcal{X}$, vrijedi $0 \leq D(\mathbb{P}_X \parallel \mathbb{P}_Y) = \log|\mathcal{X}| - H(X) = H(Y) - H(X)$. To znači da na konačnom skupu $\mathcal{X}$ uniformna distribucija ima najveću entropiju i jednaka je $\log|\mathcal{X}|$.

**Teorem 4.3** (Hanova nejednakost)**.** *Neka su $X_1, \dots, X_n$ diskretne slučajne varijable. Tada vrijedi*

$$H(\mathbf{X}) \leq \frac{1}{n-1} \sum_{i=1}^{n} H\Big(\mathbf{X}^{(-i)}\Big).$$

*Dokaz.* Vidi [4, Theorem 10]. □

**Teorem 4.4** (Hanova nejednakost za relativne entropije)**.** *Neka je $\mathcal{X}$ prebrojiv skup, neka su $P_1, \dots, P_n$ vjerojatnosti na $\mathcal{X}$, neka je $P = P_1 \times \cdots \times P_n$ i neka je $Q$ proizvoljna vjerojatnost na $\mathcal{X}^n$. Označimo sa $P^{(-i)}$, $Q^{(-i)}$ marginalne vjerojatnosti*

$$f_{Q^{(-i)}}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \sum_{x_i \in \mathcal{X}} f_Q(x_1, \dots, x_n)$$
$$f_{P^{(-i)}}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \sum_{x_i \in \mathcal{X}} f_P(x_1, \dots, x_n) = \sum_{x_i \in \mathcal{X}} f_{P_1}(x_1) \cdots f_{P_n}(x_n).$$

*Tada vrijedi*

$$D(Q \parallel P) \geq \frac{1}{n-1} \sum_{i=1}^{n} D\Big(Q^{(-i)} \,\Big\|\, P^{(-i)}\Big),$$

*ili ekvivalentno*

$$D(Q \parallel P) \leq \sum_{i=1}^{n} \Big( D(Q \parallel P) - D\Big(Q^{(-i)} \,\Big\|\, P^{(-i)}\Big)\Big). \tag{4.1}$$

*Dokaz.* Vidi [4, Theorem 11]. □

Prisjetimo se da smo u Teoremu 3.2 za $Z = g(X_1, \dots, X_n)$ pokazali da vrijedi

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}\Big[(Z - \mathbb{E}_i[Z])^2\Big],$$

što uz oznaku $\Phi(x) = x^2$ još možemo zapisati kao

$$\mathbb{E}[\Phi(Z)] - \Phi(\mathbb{E}[Z]) \leq \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}_i[\Phi(Z)] - \Phi(\mathbb{E}_i[Z])].$$

U nastavku pokazujemo da gornja nejednakost vrijedi i za $\Phi(x) = x\log(x)$. Posebno, lijevu stranu nejednakosti ćemo u tom slučaju moći zapisati kao relativnu entropiju između distribucije s gustoćom $g \cdot f_{\mathbf{X}}$ i distribucije od $\mathbf{X}$.

**Teorem 4.5.** *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable s vrijednostima u $\mathcal{X}$ i neka je $h\colon \mathcal{X}^n \to \langle 0, \infty\rangle$ Borelova. Neka je $Y = h(\mathbf{X})$ i $\Phi(x) = x\log(x)$. Ako je $\mathbb{E}[\Phi(Y)] < \infty$, vrijedi*

$$\mathbb{E}[\Phi(Y))] - \Phi(\mathbb{E}[Y]) \leq \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}_i[\Phi(Y)] - \Phi(\mathbb{E}_i[Y])]. \tag{4.2}$$

*Za nenegativnu slučajnu varijablu Z, veličinu $\mathbb{E}[\Phi(Z))] - \Phi(\mathbb{E}[Z])$ još zovemo $\Phi$-entropija od Z, a svojstvo* (4.2) *subaditivnost* $\Phi$*-entropije.*

*Dokaz.* Tvrdnju ćemo pokazati samo za diskretne slučajne varijable $X_1, \dots, X_n$, a slijedit će iz prethodnog teorema. Za potpuni dokaz tvrdnje vidi [1, Theorem 4.22]. Funkcija $\Phi$ je konveksna na $\langle 0, \infty\rangle$, pa iz Jensenove nejednakosti slijedi da su sva očekivanja konačna. Lako vidimo da za $c > 0$ vrijedi

$$\mathbb{E}[\Phi(cY)] - \Phi(\mathbb{E}[cY]) = c(\mathbb{E}[\Phi(Y)] - \Phi(\mathbb{E}[Y])),$$

a sličnu jednakost imamo i za uvjetno očekivanje $\mathbb{E}_i$. Zaključujemo da ako tvrdnja vrijedi za $Y$, vrijedit će i za $cY$, pa BSOMP $\mathbb{E}[Y] = 1$. Neka je $P = \mathbb{P}_{\mathbf{X}} = \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$. Slijedi da je mjera $Q$ na $\mathcal{X}^n$ s gustoćom $f_Q(\mathbf{x}) = h(\mathbf{x}) f_P(\mathbf{x})$ vjerojatnosna. Zato vrijedi

$$\begin{aligned}
\mathbb{E}[\Phi(Y))] - \Phi(\mathbb{E}[Y]) = \mathbb{E}[\Phi(Y))] &= \sum_{\mathbf{x}\in\mathcal{X}^n} h(\mathbf{x}) \log(h(\mathbf{x})) f_P(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in\mathcal{X}^n} f_Q(\mathbf{x}) \log\left(\frac{f_Q(\mathbf{x})}{f_P(\mathbf{x})}\right) = D(Q \parallel P),
\end{aligned}$$

pa vrijedi i

$$\mathbb{E}[\mathbb{E}_i[\Phi(Y)]] = \mathbb{E}[\Phi(Y)] = D(Q \parallel P).$$

Također imamo

$$\begin{aligned}
\mathbb{E}[\Phi(\mathbb{E}_i[Y])] &= \sum_{\mathbf{x}^{(-i)}\in\mathcal{X}^{i-1}} \Phi\left(\sum_{x_i\in\mathcal{X}} h(\mathbf{x}) \frac{f_P(\mathbf{x})}{f_{P^{(-i)}}\left(\mathbf{x}^{(-i)}\right)}\right) f_{P^{(-i)}}\left(\mathbf{x}^{(-i)}\right) \\
&= \sum_{\mathbf{x}^{(-i)}\in\mathcal{X}^{i-1}} \left(\sum_{x_i\in\mathcal{X}} h(\mathbf{x}) f_P(\mathbf{x})\right) \log\left(\sum_{x_i\in\mathbf{X}} h(\mathbf{x}) \frac{f_P(\mathbf{x})}{f_{P^{(-i)}}(\mathbf{x}^{(-i)})}\right) \\
&= \sum_{\mathbf{x}^{(-i)}\in\mathcal{X}^{i-1}} f_{Q^{(-i)}}\left(\mathbf{x}^{(-i)}\right) \log\left(\frac{f_{Q^{(-i)}}\left(\mathbf{x}^{(-i)}\right)}{f_{P^{(-i)}}(\mathbf{x}^{(-i)})}\right) \\
&= D\left(Q^{(-i)} \,\middle\|\, P^{(-i)}\right).
\end{aligned}$$

Sada iz (4.1) i dobivenih jednakosti slijedi tvrdnja. □

Ideja metode entropije za dokazivanje koncentracijskih nejednakosti je da u prethodni teorem uvrstimo slučajnu varijablu $Y = e^{sZ} > 0$ za $Z = g(\mathbf{X})$. Označimo funkciju izvodnicu momenata od $Z$ sa $M_Z(s) = \mathbb{E}\left[e^{sZ}\right]$, za koju u nastavku radi jednostavnosti pretpostavljamo da je konačna. Tada na lijevoj strani nejednakosti imamo

$$\begin{aligned} s\mathbb{E}\left[Ze^{sZ}\right] - \mathbb{E}\left[Ze^{sZ}\right]\log\left(\mathbb{E}\left[e^{sZ}\right]\right) &= sM_Z'(s) - M_Z(s)\log(M_Z(s)) \\ &= s^2 M_Z(s)\left(\frac{\frac{sM_Z'(s)}{M_Z(s)} - \log(M_Z(s))}{s^2}\right) \\ &= s^2 M_Z(s)\frac{\partial}{\partial s}\left(\frac{\log(M_Z(s))}{s}\right). \end{aligned}$$

Iz gornje jednakosti i prethodnog teorema, dobit ćemo neku diferencijalnu nejendakost sa $M_Z(s)$, iz koje ćemo pokušati dobiti gornju ogradu na $M_Z(s)$. Dalje ćemo nastaviti s Chernoffljevom metodom. Da bismo si olakšali posao, prvo ćemo dodatno ograničiti desnu stranu nejednakosti iz prethodnog toerema. O tome govore sljedeće dvije tvrdnje.

**Teorem 4.6** (Logaritamska Sobolevljeva nejednakost)**.** *Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable s vrijednostima u $\mathcal{X}$ i neka su $g\colon \mathcal{X}^n \to \langle 0, \infty\rangle$, $g_i\colon \mathcal{X}^{n-1} \to \langle 0, \infty\rangle$ Borelove za $i \in \{1, \dots, n\}$. Uz oznake $Z = g(\mathbf{X})$, $Z_i = g_i\left(\mathbf{X}^{(-i)}\right)$, $\Psi(x) = e^x - x - 1$ i $s \in \mathbb{R}$ vrijedi*

$$s\mathbb{E}\left[Ze^{sZ}\right] - \mathbb{E}\left[e^{sZ}\right]\log\left(\mathbb{E}\left[e^{sZ}\right]\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[e^{sZ}\Psi(-s(Z - Z_i))\right].$$

*Dokaz.* Vidi [4, Theorem 13]. □

**Teorem 4.7** (Simetrizirana logaritamska Sobolevljeva nejednakost)**.** *Uz oznake kao u prethodnom teoremu, neka su $X_1', \dots, X_n'$ nezavisne kopije od $X_1, \dots, X_n$ i neka je $Z_i' = g\left(\mathbf{X}^{(i)}\right)$. Tada za sve $s \in \mathbb{R}$ vrijedi*

$$s\mathbb{E}\left[Ze^{sZ}\right] - \mathbb{E}\left[e^{sZ}\right]\log\left(\mathbb{E}\left[e^{sZ}\right]\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[e^{sZ}\Psi(-s(Z - Z_i'))\right].$$

*Također, uz oznaku $\tau(x) = x(e^x - 1)$, za sve $s \in \mathbb{R}$ vrijedi*

$$s\mathbb{E}\left[Ze^{sZ}\right] - \mathbb{E}\left[e^{sZ}\right]\log\left(\mathbb{E}\left[e^{sZ}\right]\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[e^{sZ}\tau(-s(Z - Z_i'))\mathbb{1}_{\{Z > Z_i'\}}\right] \tag{4.3}$$

$$s\mathbb{E}\left[Ze^{sZ}\right] - \mathbb{E}\left[e^{sZ}\right]\log\left(\mathbb{E}\left[e^{sZ}\right]\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[e^{sZ}\tau(s(Z_i' - Z))\mathbb{1}_{\{Z < Z_i'\}}\right]. \tag{4.4}$$

*Dokaz.* Vidi [4, Theorem 14]. □

Sada možemo primijeniti metodu entropije i pokazati sljedeću tvrdnju, iz koje će slijediti eksponencijalne koncentracijske nejednakosti.

**Teorem 4.8.** *Pretpostavimo da postoji konstanta $C > 0$ takva da g.s. vrijedi*

$$\sum_{i=1}^{n}(Z - Z_i')^2 \mathbb{1}_{\{Z>Z_i'\}} \leq C.$$

*Tada za sve $t \geq 0$ vrijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq e^{-t^2/4C}.$$

*Dokaz.* Koristimo oznake kao u Teoremu 4.7. Neka je $s > 0$. Primijetimo da za $x \leq 0$ vrijedi $\tau(x) = x(e^x - 1) \leq x^2$, pa imamo

$$\begin{aligned} s\mathbb{E}\Big[Ze^{sZ}\Big] - \mathbb{E}\Big[e^{sZ}\Big]\log\Big(\mathbb{E}\Big[e^{sZ}\Big]\Big) &\overset{(4.3)}{\leq} \mathbb{E}\left[e^{sZ}\sum_{i=1}^{n}\tau(-s(Z - Z_i'))\mathbb{1}_{\{Z>Z_i'\}}\right] \\ &\leq \mathbb{E}\left[e^{sZ}\sum_{i=1}^{n}s^2(Z - Z_i')^2\mathbb{1}_{\{Z>Z_i'\}}\right] \\ &\leq Cs^2\mathbb{E}\Big[e^{sZ}\Big]. \end{aligned}$$

Označimo sa $M_Z(s) = \mathbb{E}\Big[e^{sZ}\Big]$ funkciju izvodnicu momenata od $Z$. Tada gornju nejednakost možemo zapisati kao

$$\begin{aligned} &sM_Z'(s) - M_Z(s)\log(M_Z(s)) \leq Cs^2 M_Z(s) \\ \iff &\frac{\partial}{\partial s}\left(\frac{\log(M_Z(s))}{s}\right) \leq C. \end{aligned}$$

Označimo $H(s) := \frac{\log(M_Z(s))}{s}$. Tada je $H'(s) \leq C$. Iz L'Hospitalovog pravila slijedi

$$\lim_{s\to 0+} H(s) = \lim_{s\to 0+}\frac{\log(M_Z(s))}{s} = \lim_{s\to 0+}\frac{M_Z'(s)}{M_Z(s)} = \mathbb{E}[Z].$$

Integriranjem dobijemo

$$\frac{\log(M_Z(s))}{s} = H(s) = \lim_{s\to 0+} H(s) + \int_0^s H'(t)dt \leq \mathbb{E}[Z] + Cs,$$

iz čega slijedi

$$M_Z(s) \leq e^{s\mathbb{E}[Z]+Cs^2}.$$

Još samo preostaje primijeniti Chernoffljevu metodu. Za $t > 0$ vrijedi

$$\mathbb{P}(Z \geq \mathbb{E}[Z] + t) \leq \frac{M_Z(s)}{e^{s\mathbb{E}[Z]+st}} \leq e^{Cs^2-st} \overset{s^*=\frac{t}{2C}}{\Longrightarrow} \mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq e^{-t^2/4C}. \qquad \square$$

Ako pretpostavimo da $g$ ima svojstvo omeđenih razlika s konstantama $(c_i)_{i=1}^n$, iz gornjeg teorema dobijemo McDiarmidovu nejednakost, ali s lošijom konstantom od one u Teoremu 3.9. Pokažimo jednu primjenu prethodnog teorema.

**Primjer 4.9** (Najveća svojstvena vrijednost u slučajnoj simetričnoj matrici)**.** *Neka je $A$ slučajna simetrična realna matrica, čije su vrijednosti $X_{i,j}$, $1 \leq i \leq j \leq n$ nezavisne slučajne varijable i $|X_{i,j}| \leq 1$. Neka je $Z$ najveća svojstvena vrijednost matrice $A$ i neka je $v = (v_1, \dots, v_n)$ pripadni slučajni jedinični svojstveni vektor. Provjerimo prvo da je funkcija izvodnica momenata od $Z$ konačna. To vrijedi jer je $Z$ ograničena*

$$|Z|^2 = \|Zv\|_2^2 = \|Av\|_2^2 = \sum_{i=1}^n \left( \sum_{j=1}^n X_{ij} v_j \right)^2 \leq \sum_{i=1}^n \left( \sum_{j=1}^n X_{ij}^2 \right) \left( \sum_{j=1}^n v_j^2 \right) = \sum_{i=1}^n \sum_{j=1}^n X_{ij}^2 \leq n^2.$$

*Poznato je da vrijedi*

$$Z = v^T A v = \max_{u \in \mathbb{R}^n : \|u\|_2 = 1} u^T A u.$$

*Neka je $A'_{i,j}$ slučajna simetrična realna matrica koju dobijemo iz $A$ zamjenom $X_{i,j}$ (kao i $X_{j,i}$) s nezavisnom kopijom $X'_{i,j}$. Neka je $Z'_{i,j}$ najveća svojstvena vrijednost od $A'_{i,j}$. Tada vrijedi*

$$\begin{aligned}
(Z - Z'_{i,j}) \mathbb{1}_{\{Z > Z'_{i,j}\}} &= \left( \max_{u \in \mathbb{R}^n : \|u\|_2 = 1} u^T A u - \max_{u \in \mathbb{R}^n : \|u\|_2 = 1} u^T A'_{i,j} u \right) \mathbb{1}_{\{Z > Z'_{i,j}\}} \\
&\leq \left( v^T A v - v^T A'_{i,j} v \right) \mathbb{1}_{\{Z > Z'_{i,j}\}} \\
&= \left( v^T \left( A - A'_{i,j} \right) v \right) \mathbb{1}_{\{Z > Z'_{i,j}\}} \\
&= 2 v_i v_j \left( X_{i,j} - X'_{i,j} \right) \mathbb{1}_{\{Z > Z'_{i,j}\}} \\
&\leq 4 |v_i v_j|.
\end{aligned}$$

*Zato slijedi*

$$\sum_{1 \leq i \leq j \leq n} (Z - Z'_{i,j})^2 \mathbb{1}_{\{Z > Z'_{i,j}\}} \leq 16 \sum_{1 \leq i \leq j \leq n} v_i^2 v_j^2 \leq 16 \sum_{1 \leq i, j \leq n} v_i^2 v_j^2 = 16 \left( \sum_{i=1}^n v_i^2 \right)^2 = 16.$$

*Sada iz Teorema 4.8 slijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq e^{-t^2/64}.$$

## 4.1 Samoomeđujuće funkcije

Iz prethodnih rezultata slijedit će eksponencijalne koncentracijske nejednakosti za samoomeđujuće funkcije, koje smo definirali u prethodnom poglavlju. Metodom entropije može se pokazati sljedeći teorem.

**Teorem 4.10.** *Neka funkcija g ima samoomeđujuće svojstvo i neka su*

$$h(u) = (1+u)\log(1+u) - u$$
$$\Psi(v) = e^v - v - 1,$$

*kao u Teoremima 2.11 i 4.6.*

*(i) Za svaki $s \in \mathbb{R}$ vrijedi*

$$\log\left(\mathbb{E}\left[e^{s(Z-\mathbb{E}[Z])}\right]\right) \leq \mathbb{E}[Z]\Psi(s).$$

*(ii) Za svaki $t > 0$ vrijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq \exp\left(-\mathbb{E}[Z]h\left(\frac{t}{\mathbb{E}[Z]}\right)\right) \leq \exp\left(-\frac{t^2}{2\mathbb{E}[Z] + \frac{2}{3}t}\right).$$

*(iii) Za svaki $0 < t \leq \mathbb{E}[Z]$ vrijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \leq -t) \leq \exp\left(-\mathbb{E}[Z]h\left(-\frac{t}{\mathbb{E}[Z]}\right)\right) \leq \exp\left(-\frac{t^2}{2\mathbb{E}[Z]}\right).$$

*Dokaz.* Vidi [4, Theorem 16]. □

Gornje nejednakosti imaju sličan oblik kao Bennetova i Bernsteinova nejednakost, a glavna razlika je da smo Var(Z) zamijenili gornjom ogradom $\mathbb{E}[Z]$. U nastavku definiramo novu klasu samoomeđujućih funkcija koje zovemo kombinatorne entropije.

**Definicija 4.11** (Kombinatorna entropija)**.** *Neka su $\mathcal{X}$ i $\mathcal{Y}$ skupovi, $|\mathcal{Y}| < \infty$ i neka je* $\mathrm{tr}\colon \cup_{n=1}^{\infty} \mathcal{X}^n \to \cup_{n=1}^{\infty}\mathcal{P}(\mathcal{Y}^n)$ *neko preslikavanje koje svakom $\mathbf{x} \in \mathcal{X}^n$ pridružuje skup* $\mathrm{tr}(\mathbf{x}) \subseteq \mathcal{Y}^n$*. Pretpostavimo da za svaki $n \in \mathbb{N}$ i svaki $\mathbf{x} \in \mathcal{X}^n$ vrijedi*

$$\mathrm{tr}\left(\mathbf{x}^{(-i)}\right) = \left\{\mathbf{y}^{(-i)} : \mathbf{y} \in \mathrm{tr}(\mathbf{x})\right\}.$$

*Pripadnu kombinatornu entropiju definiramo kao $h(\mathbf{x}) = \log_b|\mathrm{tr}(\mathbf{x})|$, za $b > 0$.*

**Lema 4.12.** *Neka je $h_n = h|_{\mathcal{X}^n} : \mathcal{X}^n \to \mathcal{P}(\mathcal{Y}^n)$ restrikcija kombinatorne entropije na $\mathcal{X}^n$. Ako vrijedi*

$$h_n(\mathbf{x}) - h_{n-1}\left(\mathbf{x}^{(-i)}\right) \leq 1,$$

*onda $h_n$ ima samoomeđujuće svojstvo.*

*Dokaz.* Neka je $\mathbf{x} \in \mathcal{X}^n$ proizvoljan. Iz definicije kombinatorne entropije slijedi $h_{n-1}\left(\mathbf{x}^{(-i)}\right) \leq h_n(\mathbf{x})$. Dovoljno je još pokazati da vrijedi

$$\sum_{i=1}^{n}\left(h_n(\mathbf{x}) - h_{n-1}\left(\mathbf{x}^{(-i)}\right)\right) \leq h_n(\mathbf{x}). \tag{4.5}$$

Neka je $\mathbf{Y} = (Y_1, \dots, Y_n) \in \mathcal{Y}^n$ slučajan vektor s uniformnom distribucijom na $\mathrm{tr}(\mathbf{x})$. Znamo da je

$$h_n(\mathbf{x}) = \frac{\log(|\mathrm{tr}(\mathbf{x})|)}{\log(b)} = \frac{H(\mathbf{Y})}{\log(b)}.$$

Neka je $\mathbf{Z}^{(-i)} \in \mathcal{Y}^{n-1}$ slučajan vektor s uniformnom distribucijom na $\mathrm{tr}\left(\mathbf{x}^{(-i)}\right)$. Kako i $\mathbf{Y}^{(-i)}$ poprima vrijednosti u $\mathrm{tr}\left(\mathbf{x}^{(-i)}\right)$, a uniformna distribucija maksimizira entropiju, slijedi

$$h_{n-1}\left(\mathbf{x}^{(-i)}\right) = \frac{H\left(\mathbf{Z}^{(-i)}\right)}{\log(b)} \geq \frac{H\left(\mathbf{Y}^{(-i)}\right)}{\log(b)}.$$

Iz Hanove nejednakost slijedi

$$h_n(\mathbf{x}) = \frac{H(\mathbf{Y})}{\log(b)} \leq \frac{1}{n-1}\sum_{i=1}^{n}\frac{H\left(\mathbf{Y}^{(-i)}\right)}{\log(b)} \leq \frac{1}{n-1}\sum_{i=1}^{n} h_{n-1}\left(\mathbf{x}^{(-i)}\right),$$

a to je upravo (4.5). □

Primijetimo da je uvjet leme uvijek zadovoljen za $b = |Y| < \infty$, jer vrijedi

$$\mathrm{tr}(\mathbf{x}) \subseteq \left\{(y_1, \dots, y_n) \colon (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) \in \mathrm{tr}\left(\mathbf{x}^{(-i)}\right), y_i \in \mathcal{Y}\right\},$$

pa je $|\mathrm{tr}(\mathbf{x})| \leq |Y|\left|\mathrm{tr}\left(\mathbf{x}^{(-i)}\right)\right| \implies \log_{|Y|}|\mathrm{tr}(\mathbf{x})| - \log_{|Y|}\left|\mathrm{tr}\left(\mathbf{x}^{(-i)}\right)\right| \leq 1$. Iz Teorema 4.10 lagano dobijemo sljedeći teorem.

**Teorem 4.13.** *Neka je $h(\mathbf{x}) = \log_b|\mathrm{tr}(\mathbf{x})|$ kombinatorna entropija takva da za svaki $\mathbf{x} \in \mathcal{X}^n$ vrijedi*

$$h(\mathbf{x}) - h\left(\mathbf{x}^{(-i)}\right) \leq 1, \quad \textit{za svaki } i = 1, \dots, n.$$

*Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable s vrijednostima u $\mathcal{X}$ i $Z = h(\mathbf{X})$.*

*(i) Za svaki $t > 0$ vrijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq \exp\left(-\frac{t^2}{2\mathbb{E}[Z] + \frac{2}{3}t}\right).$$

*(ii) Za svaki* $0 < t \leq \mathbb{E}[Z]$ *vrijedi*

$$\mathbb{P}(Z - \mathbb{E}[Z] \leq -t) \leq \exp\left(-\frac{t^2}{2\mathbb{E}[Z]}\right).$$

*(iii) Također vrijedi*

$$\mathbb{E}[\log_b |\mathrm{tr}(\mathbf{X})|] \leq \log_b(\mathbb{E}[|\mathrm{tr}(\mathbf{X})|]) \leq \frac{b-1}{\log(b)} \mathbb{E}[\log_b |\mathrm{tr}(\mathbf{X})|].$$

*Dokaz.* Iz prethodne leme slijedi da $h$ ima samoomeđujuće svojstvo, pa iz Teorema 4.10 odmah dobijemo prve dvije nejednakosti. Iz istog teorema za $s = \log(b)$ slijedi

$$\log\left(\mathbb{E}\left[b^{(Z-\mathbb{E}[Z])}\right]\right) \leq \mathbb{E}[Z](b - \log(b) - 1) \iff \log_b(\mathbb{E}[|\mathrm{tr}(\mathbf{X})|]) \leq \frac{b-1}{\log(b)} \mathbb{E}[\log_b |\mathrm{tr}(\mathbf{X})|],$$

dok nejednakost $\mathbb{E}[\log_b |\mathrm{tr}(\mathbf{X})|] \leq \log_b(\mathbb{E}[|\mathrm{tr}(\mathbf{X})|])$ slijedi iz Jensenove nejednakosti. □

Tipičan primjer kombinatorne entropije s kojom se susrećemo u teoriji statističkog učenja je VC entropija.

**Primjer 4.14** (VC entropija)**.** *Neka je* $\mathcal{P}(X)$ *skup,* $\mathcal{Y} = \{0, 1\}$ *i neka je* $\mathcal{A} \subseteq \mathcal{P}(\mathcal{X})$*. Definirajmo preslikavanje* $\mathrm{tr}\colon \cup_{n=1}^{\infty} \mathcal{X}^n \to \cup_{n=1}^{\infty} \mathcal{P}(\mathcal{Y}^n)$

$$\mathrm{tr}(x_1, \ldots, x_n) = \{(\mathbb{1}_A(x_1), \ldots, \mathbb{1}_A(x_n))\colon A \in \mathcal{A}\} \subseteq \mathcal{Y}^n.$$

*Neka je* $T(\mathbf{x}) = |\mathrm{tr}(\mathbf{x})|$*. Jasno je da vrijedi* $\mathrm{tr}\left(\mathbf{x}^{(-i)}\right) = \left\{\mathbf{y}^{(-i)}\colon \mathbf{y} \in \mathrm{tr}(\mathbf{x})\right\}$*, pa je* $h(\mathbf{x}) = \log_2(T(\mathbf{x}))$ *kombinatorna entropija. Kako je*

$$|\{(\mathbb{1}_A(x_1), \ldots, \mathbb{1}_A(x_n))\colon A \in \mathcal{A}\}| = |\{A \cap \{x_1, \ldots, x_n\}\colon A \in \mathcal{A}\}|,$$

*vidimo da je gore definirana funkcija* $T(\mathbf{x})$ *jednaka funkciji* $T(\mathbf{x})$ *iz Primjera 3.19. Funkciju* $h(\mathbf{x}) = \log_2(T(\mathbf{x}))$ *zovemo VC entropija, a slučajnu varijablu* $Z = h(\mathbf{X})$ *zovemo slučajna VC entropija. Slučajna VC entropija pojavljuje se u brojnim rezultatima u teoriji statističkog učenja. Slična je VC dimenziji, ali daje preciznije rezultate jer ovisi o distribuciji podataka. Uzmemo li* $b = |Y| = 2$*, možemo primijeniti prethodni teorem i dobijemo da slučajna VC entropija* $Z = h(\mathbf{X})$ *zadovoljava navedene eksponencijalne koncentracijske nejednakosti. Također vrijedi*

$$\mathbb{E}[\log_2 T(\mathbf{X})] \leq \log_2(\mathbb{E}[T(\mathbf{X})]) \leq \frac{1}{\log(2)} \mathbb{E}[\log_2 T(\mathbf{X})],$$

*što pokazuje da su očekivana slučajna VC entropija i veličina* $\log_2(\mathbb{E}[|\mathrm{tr}(\mathbf{X})|])$ *usko povezane. Pomoću tih veličina mogu se dati nužni i dovoljni uvjeti za konzistentnu i brzu konvergenciju minimizacije empirijskog rizika. Za definicije i detalje vidi [6, Section 2.7].*

# Poglavlje 5

# Supremum slučajnog procesa

U ovom poglavlju pokazat ćemo koncentracijske nejednakosti za supremum slučajnog procesa $\sup_{t\in T} X_t$, konkretnije pokazat ćemo ograde na očekivanje supremuma. Glavni rezultat će biti Dudleyeva nejednakost, koja će imati niz korisnih posljedica iz teorije statističkog učenja, od kojih smo neke već spomenuli kroz primjere. Kada radimo sa supremumom, uvijek se otvara pitanje izmjerivosti. Sljedeća napomena daje jednostavno rješenje kada radimo s očekivanjem supremuma.

**Napomena 5.1.** *Neka je $(X_t)_{t\in T}$ neki slučajni proces na vjerojatnosnom prostoru $(\Omega, \Sigma, \mathbb{P})$. Da bismo izbjegli probleme s izmjerivosti kod računanja supremuma očekivanja $\mathbb{E}[\sup_{t\in T} X_t]$, uvodimo oznaku*

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] := \sup_{\substack{S\subseteq T\\ |S|<\infty}} \mathbb{E}\left[\max_{t\in S} X_t\right]. \tag{5.1}$$

Gornja oznaka je puno praktičnija, a opet često vrijedi $\mathbf{E}[\sup_{t\in T} X_t] = \mathbb{E}[\sup_{t\in T} X_t]$. Navedimo nekoliko primjera kada je to slučaj.

(i) Ako $T$ je konačan, izrazi su očito jednaki.

(ii) Ako je $T$ prebrojiv i $\sup_{t\in T}|X_t|$ integrabilna, izrazi će biti jednaki po Lebesgueovom teoremu o dominiranoj konvergenciji.

(iii) Ako je metrički prostor $\big(\{X_t \colon t \in T\}, \|\cdot\|_{L^\infty(\mathbb{P})}\big)$ separabilan, supremum će biti jednak supremumu po gustom prebrojivom podskupu, pa opet ako je $\sup_{t\in T}|X_t|$ integrabilna vrijedi jednakost.

Osim gore navedenih primjera, uvedena oznaka nam omogućava da radimo s konačnim podfamilijama. O tome govori sljedeća napomena.

**Napomena 5.2.** *Pretpostavimo da za neku familiju $\mathcal{F}$ želimo pokazati nejednakost ovog oblika*

$$\mathbf{E}\left[\sup_{f\in\mathcal{F}} X_f\right] \leq C \cdot \mathrm{RHS}(\mathcal{F}), \tag{5.2}$$

*za neku apsolutnu konstantu $C > 0$. Ako za desnu stranu nejednakosti vrijedi*

$$\mathrm{RHS}(\mathcal{G}) \leq D \cdot \mathrm{RHS}(\mathcal{F}) \qquad \forall \mathcal{G} \subseteq \mathcal{F}, |\mathcal{G}| < \infty, \tag{5.3}$$

*za neku apsolutnu konstantu $D > 0$, bit će dovoljno nejednakost* (5.2) *pokazati za sve konačne podfamilije, odnosno da vrijedi*

$$\mathbb{E}\left[\max_{g\in\mathcal{G}} \left|X_g\right|\right] \leq C \cdot \mathrm{RHS}(\mathcal{G}), \qquad \forall \mathcal{G} \subseteq \mathcal{F}, |\mathcal{G}| < \infty. \tag{5.4}$$

*Ako pokažemo da gornja nejednakost vrijedi za sve konačne podfamilije, vrijedit će*

$$\mathbb{E}\left[\max_{g\in\mathcal{G}} \left|X_g\right|\right] \overset{(5.4)}{\leq} C \cdot \mathrm{RHS}(\mathcal{G}) \overset{(5.3)}{\leq} CD \cdot \mathrm{RHS}(\mathcal{F}).$$

*Računanjem supremuma po svim $\mathcal{G} \subseteq \mathcal{F}$, $|\mathcal{G}| < \infty$ slijedit će nejednakost* (5.2) *s apsolutnom konstantom $CD$.*

Pridružimo li slučajnom procesu $(X_t)_{t\in T}$ metriku, njegovu kompleksnost možemo mjeriti preko brojeva natkrivanja.

**Definicija 5.3.** *Neka je $(T, d)$ metrički prostor. Neka je $K \subseteq T$ i neka je $\varepsilon > 0$. Za podskup $N \subseteq K$ kažemo da je $\varepsilon$-mreža za $K$ ako za svaki $x \in K$ postoji $y \in N$ takav da je $d(x, y) \leq \varepsilon$. Minimalni kardinalitet $\varepsilon$-mreže za $K$ zovemo $\varepsilon$-broj natkrivanja i označavamo ga sa $\mathcal{N}(K, d, \varepsilon)$.*

Vrijedi sljedeća monotonost brojeva natkrivanja.

**Propozicija 5.4** (Monotonost brojeva natkrivanja)**.** *Neka je $(T, d)$ metrički prostor.*

*(i) Ako su $0 < \varepsilon_1 \leq \varepsilon_2$, tada vrijedi*

$$\mathcal{N}(K, d, \varepsilon_1) \geq \mathcal{N}(K, d, \varepsilon_2). \tag{5.5}$$

*(ii) Ako su $L \subseteq K \subseteq T$ i $\varepsilon > 0$, tada vrijedi*

$$\mathcal{N}(L, d, 2\varepsilon) \leq \mathcal{N}(K, d, \varepsilon). \tag{5.6}$$

*Dokaz.* Nejednakost (5.5) je očita, jer je svaki $\varepsilon_1$-mreža za $K$ ujedno i $\varepsilon_2$-mreža za $K$. Pokažimo nejednakost (5.6). Neka je $\varepsilon > 0$ i neka je $\mathcal{K}$ minimalni $\varepsilon$-mreža za $K$, odnosno takav da vrijedi $|\mathcal{K}| = \mathcal{N}(K, d, \varepsilon)$. Tada je $L \subseteq K \subseteq \cup_{k \in \mathcal{K}} \overline{K}(k, \varepsilon)$. Definirajmo $\mathcal{K}' := \left\{k \in \mathcal{K} \colon L \cap \overline{K}(k, \varepsilon) \neq \emptyset\right\}$. Tada opet vrijedi $L \subseteq \cup_{k \in \mathcal{K}'} \overline{K}(k, \varepsilon)$. Za svaki $k \in \mathcal{K}'$, označimo sa $l_k$ proizvoljan element skupa $L \cap \overline{K}(k, \varepsilon)$ i definirajmo $\mathcal{L} = \{l_k, k \in \mathcal{K}'\} \subseteq L$. Tada po nejednakosti trokuta vrijedi

$$L \subseteq \bigcup_{k \in \mathcal{K}'} \underbrace{\overline{K}(k, \varepsilon)}_{\subseteq \overline{K}(l_k, 2\varepsilon)} \subseteq \bigcup_{l \in \mathcal{L}} \overline{K}(l, 2\varepsilon).$$

Dakle, $\mathcal{L}$ je $(2\varepsilon)$-mreža za $L$, pa vrijedi

$$\mathcal{N}(L, d, 2\varepsilon) \leq |\mathcal{L}| \leq |\mathcal{K}'| \leq |\mathcal{K}| = \mathcal{N}(K, d, \varepsilon). \qquad \square$$

Sudakovljeva nejednakost daje donju ogradu na očekivanje supremuma gaussovskog procesa preko brojeva natrkivanja.

**Definicija 5.5** (Gaussovski proces)**.** *Kažemo da je slučajni $(X_t)_{t \in T}$ proces gaussovski ako je, za svaki konačan podskup $S \subseteq T$, slučajni vektor $(X_t)_{t \in S}$ normalno distribuiran.*

**Teorem 5.6** (Sudakov)**.** *Neka je $(X_t)_{t \in T}$ Gaussovksi slučajani proces s očekivanjem nula. Tada vrijedi*

$$\mathbf{E}\left[\sup_{t \in T} X_t\right] \gtrsim \sup_{\varepsilon \geq 0}\Big(\varepsilon \sqrt{\log \mathcal{N}(T, d, \varepsilon)}\Big),$$

*gdje je $d$ metrika na $T$ zadana sa*

$$d(s,t) := \|X_s - X_t\|_{L^2(\mathbb{P})} = \sqrt{\mathbb{E}\Big[(X_s - X_t)^2\Big]}.$$

*Dokaz.* Vidi [7, Theorem 7.4.1]. $\square$

Veličinu $\log \mathcal{N}(T, d, \varepsilon)$ iz gornjeg teorema još zovemo metrička entropija od $T$. U nastavku dajemo sličnu gornju ogradu za slučajne procese sa subgaussovskim prirastima.

**Definicija 5.7** (Subgaussovski prirasti)**.** *Neka je $(X_t)_{t \in T}$ slučajni proces na metričkom prostoru $(T, d)$. Kažemo da proces ima subgaussovske priraste ako postoji konstanta $K_d \geq 0$ takav da vrijedi*

$$\|X_s - X_t\|_{\psi_2} \leq K_d \cdot d(s,t) \quad \text{za sve } s, t \in T. \tag{5.7}$$

Procesi sa subgaussovskim prirastima su generalizacija gaussovskih procesa, što pokazuje sljedeći primjer.

**Primjer 5.8.** *Neka je $(X_t)_{t\in T}$ gaussovski proces s konstantnim očekivanjem. Definirajmo metriku na $T$ sa*

$$d(s,t) = \|X_s - X_t\|_{L^2}, \quad s,t \in T.$$

*Očito je $X_s - X_t$ normalno distribuira za svake $s,t \in T$. U Primjeru 2.16 (c) vidjeli smo da vrijedi $\|X_s - X_t\|_{\psi_2} \lesssim \|X_s - X_t\|_{L^2(\mathbb{P})}$, pa je $(X_t)_{t\in T}$ subgaussovski proces.*

Dudleyeva nejednakost daje gornju ogradu za očekivanje supremuma subgaussovskog procesa preko metričke entropije. Metodu kojom se pokazuje zovemo metoda ulančavanja (eng. chaining), a ideja je da slučajni proces $(X_t)_{t\in T}$ aproksimiramo sve finijim $\varepsilon_k$-mrežama. Svaki $X_t$ aproksimirat ćemo nizom $\left(X_{\pi_k(t)}\right)_k$ aproksimacija iz $\varepsilon_k$-mreža, a zatim ćemo naći gornje ograde za razlike između uzastopnih aproksimacija $X_{\pi_k(t)} - X_{\pi_{k+1}(t)}$. Zbrajanjem tih ograda dobit ćemo gornju ogradu na supremum procesa.

**Teorem 5.9** (Dudley, diskretna verzija)**.** *Neka je $(X_t)_{t\in T}$ slučajni proces na metričkom prostoru $(T,d)$ s očekivanjem* 0 *i subgaussovskim prirastima, kao u* (5.7)*. Tada vrijedi*

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \lesssim K_d \sum_{k\in\mathbb{Z}} 2^{-k} \sqrt{\log \mathcal{N}(T,d,2^{-k})}.$$

*Dokaz.* Primijetimo da za $S \subseteq T$, $|S| < \infty$ vrijedi

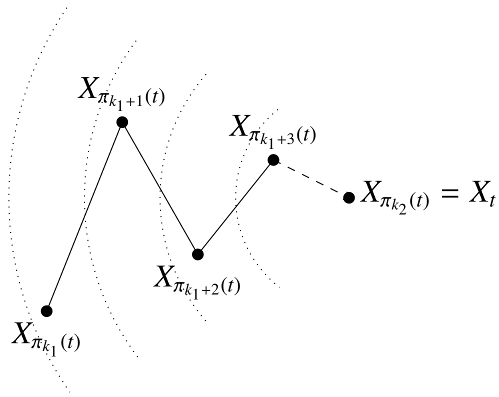
$$\begin{aligned} K_d \sum_{k\in\mathbb{Z}} 2^{-k} \sqrt{\log \mathcal{N}(S,d,2^{-k})} &\overset{(5.6)}{\leq} K_d \sum_{k\in\mathbb{Z}} 2^{-k} \sqrt{\log \mathcal{N}(T,d,2^{-k-1})} \\ &= 2K_d \sum_{k\in\mathbb{Z}} 2^{-k} \sqrt{\log \mathcal{N}(T,d,2^{-k})}, \end{aligned}$$

pa po Napomeni 5.2 možemo bez smanjenja općenitosti pretpostaviti da je $T$ konačan.

Definirajmo niz $(\varepsilon_k)_{k\in\mathbb{Z}}$ sa $\varepsilon_k = 2^{-k}$ i označimo sa $T_k$ minimalnu $\varepsilon_k$-mrežu za $T$ za svaki $k \in \mathbb{Z}$, odnosno takvu da vrijedi $|T_k| = \mathcal{N}(T,d,\varepsilon_k)$. Tada je $(|T_k|)_{k\in Z}$ neopadajući niz. Kako je $T$ konačan, postojat će $k \in \mathbb{Z}$ takav da je $T_k$ jednočlan. Označimo sa $k_1$ najveći takav $k$ i sa $t_0$ jedini element skupa $T_{k_1}$. Također, postoji $k_2 \in \mathbb{Z}$ takav da je $T_{k_2} = T$.

Za svaki $t \in T$, označimo sa $\pi_k(t)$ njoj (bilo koju) najbližu točku iz $T_k$. Tada vrijedi

$$d(t,\pi_k(t)) \leq \varepsilon_k.$$




Slika 5.1: Svaki $X_t$ aproksimiramo nizom sve finijih aproksimacija $X_{\pi_k(t)} \in \overline{K}(X_t, \varepsilon_k)$.

Kako je $\mathbb{E}[X_{t_0}] = 0$, imamo

$$\mathbb{E}\Big[\max_{t\in T} X_t\Big] = \mathbb{E}\Big[\max_{t\in T} X_t\Big] - \mathbb{E}[X_{t_0}] = \mathbb{E}\Big[\max_{t\in T} X_t - X_{t_0}\Big] = \mathbb{E}\Big[\max_{t\in T}(X_t - X_{t_0})\Big].$$

Primijetimo da za svaki $t \in T$ vrijedi $\pi_{k_1}(t) = t_0$ i $\pi_{k_2}(t) = t$. Zato vrijedi

$$\mathbb{E}\Big[\max_{t\in T}(X_t - X_{t_0})\Big] = \mathbb{E}\left[\max_{t\in T}\left(\sum_{k=k_1+1}^{k_2} X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right)\right] \le \sum_{k=k_1+1}^{k_2} \mathbb{E}\Big[\max_{t\in T}(X_{\pi_k(t)} - X_{\pi_{k-1}(t)})\Big].$$

Primijetimo da maksimum u gornjim sumandima možemo drukčije zapisati

$$\max_{t\in T}(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}) = \max_{(s,t)\in P_k}(X_s - X_t), \tag{5.8}$$

gdje je

$$P_k \coloneqq \{(\pi_k(t), \pi_{k-1}(t)) \colon t \in T\}.$$

Za svaki $k > k_1$ je $|T_k| \ge 2$, pa vrijedi $|P_k| \ge |T_k| \ge 2$. Također imamo i gornju ogradu

$$|P_k| \le |T_k| \cdot |T_{k-1}| \le |T_k|^2 = \mathcal{N}(T, d, \varepsilon_k)^2. \tag{5.9}$$

Za svaki $t \in T$ vrijedi

$$\begin{aligned} \left\| X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \right\|_{\psi_2} &\le K_d d(\pi_k(t), \pi_{k-1}(t)) \\ &\le K_d\Big[d(\pi_k(t), t) + d(t, \pi_{k-1}(t))\Big] \\ &\le K_d(\varepsilon_k + \varepsilon_{k-1}) \\ &\le 2K_d\varepsilon_{k-1}. \end{aligned} \tag{5.10}$$

Sada za $k > k_1$ možemo iskoristiti Lemu 2.19 i dobivamo

$$\begin{aligned}
\mathbb{E}\left[\max_{t\in T}(X_{\pi_k(t)} - X_{\pi_{k-1}(t)})\right] &\overset{(5.8)}{=} \mathbb{E}\left[\max_{(s,t)\in P_k}(X_s - X_t)\right] \\
&\le \mathbb{E}\left[\max_{(s,t)\in P_k}|X_s - X_t|\right] \\
&\overset{2.19}{\le} C \max_{(s,t)\in P_k} \|X_s - X_t\|_{\psi_2} \sqrt{\log|P_k|} \\
&\overset{(5.9),(5.10)}{\le} 2\sqrt{2}CK_d\varepsilon_{k-1}\sqrt{\log \mathcal{N}(T,d,\varepsilon_k)}. \qquad (5.11)
\end{aligned}$$

Preostaje još iskoristiti sve pokazane ograde

$$\begin{aligned}
\mathbb{E}\left[\max_{t\in T} X_t\right] &\le \sum_{k=k_1+1}^{k_2} \mathbb{E}\left[\max_{t\in T}(X_{\pi_k(t)} - X_{\pi_{k-1}(t)})\right] \\
&\overset{(5.11)}{\le} 2\sqrt{2}CK_d \sum_{k=k_1+1}^{k_2} \varepsilon_{k-1}\sqrt{\log \mathcal{N}(T,d,\varepsilon_k)} \\
&= 4\sqrt{2}CK_d \sum_{k=k_1+1}^{k_2} 2^{-k}\sqrt{\log \mathcal{N}(T,d,2^{-k})} \\
&\le (4\sqrt{2}C)K_d \sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log \mathcal{N}(T,d,2^{-k})}. \qquad \square
\end{aligned}$$

Sumu u diskretnoj verziji Dudleyeve nejednakosti lako možemo zamijeniti integralom i tako dobijemo sljedeći teorem.

**Teorem 5.10** (Dudley, integralna verzija)**.** *Neka je* $(X_t)_{t\in T}$ *slučajni proces na metričkom prostoru* $(T,d)$ *s očekivanjem* 0 *i subgaussovskim prirastima, kao u* (5.7)*. Tada vrijedi*

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \lesssim K_d \int_0^\infty \sqrt{\log \mathcal{N}(T,d,\varepsilon)}\,d\lambda(\varepsilon).$$

*Dokaz.* Slijedi iz diskretne verzije nejednakosti

$$\begin{aligned}
\mathbf{E}\Big[\sup_{t\in T} X_t\Big] &\leq CK_d \sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log \mathcal{N}(T,d,2^{-k})} \\
&= 2CK_d \sum_{k\in\mathbb{Z}} \int_{\langle 2^{-k-1},2^{-k}]} \sqrt{\log \mathcal{N}(T,d,2^{-k})}d\lambda(\varepsilon) \\
&\overset{(5.5)}{\leq} 2CK_d \sum_{k\in\mathbb{Z}} \int_{\langle 2^{-k-1},2^{-k}]} \sqrt{\log \mathcal{N}(T,d,\varepsilon)}d\lambda(\varepsilon) \\
&\overset{\text{LTMK}}{=} (2C)K_d \int_{\langle 0,\infty\rangle} \sqrt{\log \mathcal{N}(T,d,\varepsilon)}d\lambda(\varepsilon). \qquad \square
\end{aligned}$$

**Napomena 5.11.** *Iz dokaza vidimo da bez pretpostavke o očekivanju nula vrijedi nejednakost*

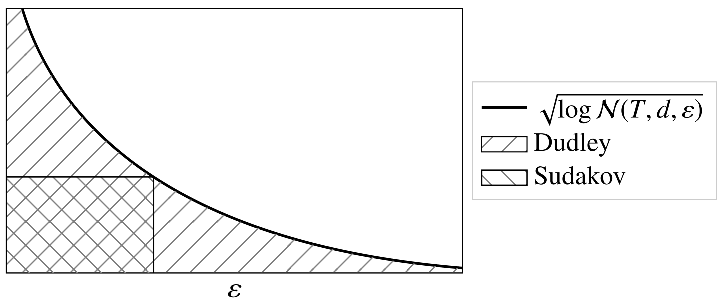
$$\mathbf{E}\Big[\sup_{t\in T}\big|X_t - X_{t_0}\big|\Big] \lesssim K_d \int_0^{\operatorname{diam}(T)} \sqrt{\log \mathcal{N}(T,d,\varepsilon)}d\lambda(\varepsilon),$$

*za proizvoljan i fiksan* $t_0 \in T$*, jer je* $\mathcal{N}(T,d,\varepsilon) = 1$ *za* $\varepsilon \geq \operatorname{diam}(T)$*. Zato vrijedi i*

$$\begin{aligned}
\mathbf{E}\Big[\sup_{s,t\in T}|X_t - X_s|\Big] &\leq \mathbf{E}\Big[\sup_{s,t\in T}\Big(\big|X_t - X_{t_0}\big| + \big|X_s - X_{t_0}\big|\Big)\Big] \leq \mathbf{E}\Big[\sup_{t\in T}\big|X_t - X_{t_0}\big|\Big] + \mathbf{E}\Big[\sup_{s\in T}\big|X_s - X_{t_0}\big|\Big] \\
&\lesssim K_d \int_0^{\operatorname{diam}(T)} \sqrt{\log \mathcal{N}(T,d,\varepsilon)}d\lambda(\varepsilon).
\end{aligned}$$

Usporedimo ograde iz Sudakovljeve i Dudleyeve nejednakosti za gaussovski proces $(X_t)_{t\in T}$ i $d(s,t) = \|X_s - X_t\|_{L^2(\mathbb{P})}$

$$\sup_{\varepsilon\geq 0}\Big(\varepsilon\sqrt{\log \mathcal{N}(T,d,\varepsilon)}\Big) \lesssim \mathbf{E}\Big[\sup_{t\in T} X_t\Big] \lesssim \int_0^\infty \sqrt{\log \mathcal{N}(T,d,\varepsilon)}d\lambda(\varepsilon).$$




Slika 5.2: Usporedba ograda iz Sudakovljeve i Dudleyeve nejednakosti.

Sudakovljeva nejednakost daje donju ogradu preko površine najvećeg pravokutnika kojeg možemo smjestiti ispod grafa funkcije $\sqrt{\log \mathcal{N}(T, d, \varepsilon)}$, dok Dudleyeva nejednakost daje gornju ogradu preko površine ispod grafa. Ako je $\mathbf{X} = (X_1, \ldots, X_n) \sim N_n(0, I_n)$ i $T \subseteq \mathbb{R}^n$, za *kanonski* gaussovski proces $X_\mathbf{t} = \langle \mathbf{X}, \mathbf{t} \rangle, \mathbf{t} \in T$ vrijedi dvostrana Sudakovljeva nejednakost

$$\sup_{\varepsilon \geq 0} \varepsilon \sqrt{\log \mathcal{N}(T, d, \varepsilon)} \lesssim \mathbf{E}\left[\sup_{\mathbf{t} \in T} X_\mathbf{t}\right] \lesssim \log(n) \cdot \sup_{\varepsilon \geq 0} \varepsilon \sqrt{\log \mathcal{N}(T, d, \varepsilon)},$$

gdje je $d(\mathbf{s}, \mathbf{t}) = \|X_\mathbf{s} - X_\mathbf{t}\|_{L^2(\mathbb{P})} = \|\mathbf{s} - \mathbf{t}\|_2$. Za dokaz vidi [7, Theorem 8.1.13]. Dakle, Sudakovljeva nejednakost za kanonski gaussovksi proces je "točna" do na faktor $\log(n)$. Sljedeći primjer pokazuje kako Dudleyeva nejednakost nije uvijek oštra.

**Primjer 5.12.** *Neka je $e_1, \ldots, e_n$ kanonska baza za $\mathbb{R}^n$. Promotrimo kanonski gaussovksi proces $(X_\mathbf{t})_{\mathbf{t} \in T}$ zadan sa*

$$T = \left\{ \frac{e_i}{\sqrt{1 + \log(i)}}, i = 1, \ldots, n \right\}.$$

*Iz Leme 2.19 i Primjera 2.16(c) slijedi*

$$\mathbb{E}\left[\max_{\mathbf{t} \in T} X_\mathbf{t}\right] = \mathbb{E}\left[\max_{i \leq n} \frac{|X_i|}{\sqrt{1 + \log(i)}}\right] \leq C,$$

*gdje je $C > 0$ apsolutna konstanta. Lako vidimo da za $2 \leq m \leq n$, $i < j \leq m$ vrijedi*

$$d\left(\frac{e_i}{\sqrt{1 + \log(i)}}, \frac{e_j}{\sqrt{1 + \log(j)}}\right) = \sqrt{\frac{1}{1 + \log(i)} + \frac{1}{1 + \log(j)}} > \frac{1}{\sqrt{\log(m)}},$$

*što znači da za $\varepsilon \leq \frac{1}{\sqrt{\log(m)}}$ vrijedi $\mathcal{N}(T, d, \varepsilon) \geq m$. Uvrstimo li $m = 2^{k^2}$, dobijemo da za $1 \leq k \leq \sqrt{\log_2(n)}$ i $\varepsilon \leq \frac{1}{\log(2)k}$ vrijedi $\mathcal{N}(T, d, \varepsilon) \geq 2^{k^2}$. Sada računamo*

$$\begin{aligned}
\int_0^\infty \sqrt{\log \mathcal{N}(T, d, \varepsilon)} d\lambda(\varepsilon) &\geq \sum_{k=1}^{\lfloor \sqrt{\log_2(n)} \rfloor} \int_{\frac{1}{\log(2)(k+1)}}^{\frac{1}{\log(2)k}} \sqrt{\log \mathcal{N}(T, d, \varepsilon)} d\lambda(\varepsilon) \\
&\geq \sum_{k=1}^{\lfloor \sqrt{\log_2(n)} \rfloor} \left( \frac{1}{\log(2)k} - \frac{1}{\log(2)(k+1)} \right) \log(2)k \\
&= \sum_{k=1}^{\lfloor \sqrt{\log_2(n)} \rfloor} \frac{1}{k+1} \xrightarrow[n \to \infty]{} \infty.
\end{aligned}$$

Iako smo u prethodnom primjeru vidjeli da Dudleyeva nejednakost nije uvijek oštra, metodom ulančavanja ipak možemo dobiti oštru dvostranu ogradu za supremum. Prisjetimo se da smo u dokazu Dudleyeve nejednakosti pokazali ogradu oblika

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \lesssim \sum_{k=1}^{\infty} \varepsilon_{k-1}\sqrt{\log|T_k|}, \tag{5.12}$$

gdje je $(e_k)_{k=0}^{\infty}$ bio padajući niz pozitivnih brojeva, a $(T_k)_{k=0}^{\infty}$ su bili $\varepsilon_k$-mreže za $T$ takvi da je $|T_0| = 1$. Konkretnije, u dokazu smo zadali niz $\varepsilon_k$ i onda smo za $T_k$ uzeli najmanje $\varepsilon_k$-mreže za $T$. Drugi pristup je da prvo odaberemo neki niz $(T_k)_{k=0}^{\infty}$ i radimo s najmanjim $\varepsilon_k$. Neka je $(T_k)_{k=0}^{\infty}$ neki niz podskupova od $T$ takav da je

$$|T_0| = 1, \quad |T_k| \leq 2^{2^k}, \quad k = 1, 2, \ldots \tag{5.13}$$

Za takve nizove $(T_k)_{k=0}^{\infty}$ kažemo da su dopustivi. Definirajmo

$$\varepsilon_k = \sup_{t\in T} d(t, T_k) = \sup_{t\in T} \inf_{s\in T_k} d(t, s).$$

Za takav odabir $T_k$ i $\varepsilon_k$ ograda (5.12) postaje

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \lesssim \sum_{k=1}^{\infty} 2^{k/2} \sup_{t\in T} d(t, T_{k-1}) \lesssim \sum_{k=0}^{\infty} 2^{k/2} \sup_{t\in T} d(t, T_k).$$

Može se pokazati da dobivena suma na desnoj strani nejednakosti i dalje ne daje oštru ogradu u situaciji iz zadnjeg primjera, neovisno o izboru dopustivog niza $(T_k)_{k=0}^{\infty}$. Konkretno, za metriku $d$ iz zadnjeg primjera vrijedi

$$\inf_{(T_k)_{k=0}^{\infty} \text{ dopustiv}} \sum_{k=0}^{\infty} 2^{k/2} \sup_{t\in T} d(t, T_k) \xrightarrow[n\to\infty]{} \infty.$$

Značajna razlika nastaje ako zamijenimo poredak sume i supremuma. Definirajmo Talagrandov $\gamma_2$ funkcional kao

$$\gamma_2(T, d) = \inf_{(T_k)_{k=0}^{\infty} \text{ dopustiv}} \sup_{t\in T} \sum_{k=0}^{\infty} 2^{k/2} d(t, T_k).$$

S Talagrandovim $\gamma_2$ funkcionalom teže je računati, ali zato daje dvostranu ogradu za gaussovske procese koja je optimalna do na apsolutnu konstantu.

**Teorem 5.13** (Talagrandov teorem o majorizacijskim mjerama)**.** *Neka je $(X_t)_{t\in T}$ gaussovski proces na skupu $T$ s očekivanjem* 0*. Tada uz metriku $d(s,t) := \|X_s - X_t\|_{L^2(\mathbb{P})}$ na $T$ vrijedi*

$$\gamma_2(T, d) \lesssim \mathbf{E}\left[\sup_{t\in T} X_t\right] \lesssim \gamma_2(T, d).$$

*Dokaz.* Vidi [7, Theorem 8.6.1]. □

## 5.1 Primjene u teoriji statističkog učenja

U ovom odjeljku pokazat ćemo neke primjene Dudleyeve nejednakosti u teoriji statističkog učenja. Pokazat ćemo ograde za empirijske procese, s kojima smo se već nekoliko puta susreli kroz primjere.

**Definicija 5.14** (Empirijski proces). *Neka su $X, X_1, \dots, X_n$ nezavisne jednako distribuirane slučajne varijable s vrijednostima u $\Omega$. Neka je $\mathcal{F}$ familija izmjerivih funkcija $f\colon \Omega \to \mathbb{R}$. Tada slučajni proces $(X_f)_{f\in\mathcal{F}}$ definiran sa*

$$X_f := \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]$$

*zovemo empirijski proces indeksiran sa $\mathcal{F}$.*

Aproksimacija očekivanja $\mathbb{E}[f(X)]$ prosjekom $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$ prisutna je u brojnim primjenama, kao što je Monte Carlo metoda računanja integrala funkcije $f$. Iz Jensenove nejednakosti slijedi

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] \leq \sqrt{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)} = \sqrt{\frac{\mathrm{Var}(X_1)}{n}},$$

što znači da je očekivana greška takve aprokcimacije za zadanu funkciju $f$ najviše reda $\frac{1}{\sqrt{n}}$. Sljedeći teorem govori da ako radimo s Lipschitzovim funkcijama, možemo postići grešku istog reda veličine uniformno po svim funkcijama.

**Definicija 5.15** (Lipschitzova funkcija). *Neka su $(X, d_X)$ i $(Y, d_Y)$ metrički prostori. Za funkciju $f\colon X \to Y$ kažemo da je Lipschitzova ako postoji $L \geq 0$ takav da vrijedi*

$$d_y(f(u), f(v)) \leq L \cdot d_X(u, v) \quad \textit{za svaki } u,\ v\ \in X.$$

*Infimum svih takvih $L \geq 0$ označavamo sa $\|f\|_{\mathrm{Lip}}$.*

**Teorem 5.16** (Uniformni zakon velikih brojeva). *Naka su $X, X_1, \dots, X_n$ nezavisne jednako distribuirane slučajne varijable s vrijednostima u $[0,1]$. Neka je $L \geq 0$ i neka je*

$$\mathcal{F} := \Big\{f\colon [0,1] \to \mathbb{R}\colon \|f\|_{\mathrm{Lip}} \leq L\Big\}.$$

*Tada vrijedi*

$$\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] \lesssim \frac{L}{\sqrt{n}}.$$

*Dokaz.* Za $L = 0$ nejednakost je trivijalna. Za $L > 0$ dovoljno je pokazati nejednakost za familiju

$$\mathcal{G} := \left\{ f \colon [0,1] \to [0,1] \colon \|f\|_{\mathrm{Lip}} \le 1 \right\}.$$

Zaista, ako pokažemo nejendakost za $\mathcal{G}$, vrijedit će

$$\begin{aligned}
\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] &= L\cdot\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\underbrace{\left(\frac{f-\min f}{L}\right)}_{\in\mathcal{G}}(X_i) - \mathbb{E}\left[\left(\frac{f-\min f}{L}\right)(X)\right]\right|\right] \\
&\le L\cdot\mathbf{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X)]\right|\right] \lesssim L\cdot\frac{1}{\sqrt{n}}.
\end{aligned}$$

U nastavku pretpostavljamo $\mathcal{F} = \mathcal{G}$. Teorem ćemo pokazati primjenom Dudleyeve nejednakosti. Prvo provjeravamo da empirijski proces $X_f$ ima subgaussovske priraste. Neka su $f, g \in \mathcal{F}$ fiksne. Tada je

$$\left\|X_f - X_g\right\|_{\psi_2} = \frac{1}{n}\left\|\sum_{i=1}^{n} Z_i\right\|_{\psi_2}, \quad \text{gdje je } Z_i := (f-g)(X_i) - \mathbb{E}[(f-g)(X)].$$

Slučajne varijable $Z_1, \ldots, Z_n$ su nezavisne, jednako distribuirane, subgaussovske s očekivanjem 0, pa vrijedi

$$\left\|X_f - X_g\right\|_{\psi_2} \overset{2.17}{\lesssim} \frac{1}{n}\sqrt{\sum_{i=1}^{n}\|Z_i\|_{\psi_2}^2} \overset{2.18}{\lesssim} \frac{1}{n}\sqrt{\sum_{i=1}^{n}\|(f-g)(X_i)\|_{\psi_2}^2} \overset{(2.9)}{\lesssim} \frac{1}{\sqrt{n}}\|f-g\|_{L^\infty}.$$

Nul-funkcija je očito u $\mathcal{F}$ i u metričkom prostoru $(\mathcal{F}, \|\cdot\|_{L^\infty})$ imamo $\operatorname{diam}(\mathcal{F}) \le 1$, pa je

$$\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|X_f\right|\right] = \mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|X_f - X_0\right|\right] \overset{5.11}{\lesssim} \frac{1}{\sqrt{n}}\int_0^1 \sqrt{\log\mathcal{N}(\mathcal{F}, \|\cdot\|_{L^\infty}, \varepsilon)}\, d\lambda(\varepsilon).$$

Uz $L = 1$ možemo pokazati nejednakost

$$\mathcal{N}(\mathcal{F}, \|\cdot\|_{L^\infty}, \varepsilon) \le Ce^{\frac{D}{\varepsilon}} \quad \text{za svaki } \varepsilon \in \langle 0, 1], \tag{5.14}$$

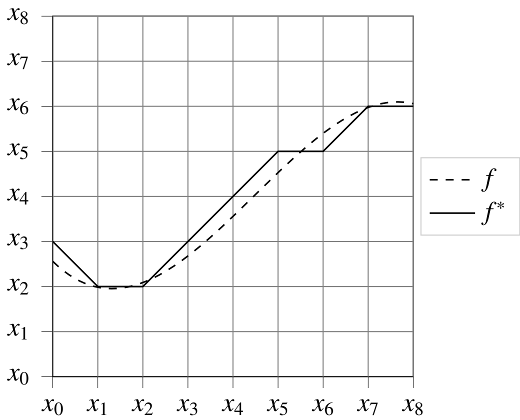
gdje su $C, D > 0$ apsolutne konstante. Iz toga će slijediti

$$\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|X_f\right|\right] \overset{(5.14)}{\lesssim} \frac{1}{\sqrt{n}}\underbrace{\int_0^1 \sqrt{\ln(C) + \frac{D}{\varepsilon}}\, d\varepsilon}_{<\infty} \lesssim \frac{1}{\sqrt{n}}.$$

I teorem je dokazan. Pokažimo još nejednakost (5.14). Neka je $\varepsilon > 0$, $m = \left\lceil \frac{1}{\varepsilon} \right\rceil$ i neka je $(x_i)_{i=0}^m$ niz definiran sa $x_i = \frac{i}{m}$. Za $f \in \mathcal{F}$, neka je $f^* \colon [0,1] \to [0,1]$ definirana sa

(i) $f^*(x_i) = \arg\min_{x_j} \left| f(x_i) - x_j \right|$, za svaki $i$

(ii) $f^*|_{[x_i, x_{i+1}]}$ je linearna za svaki $i$




Slika 5.3: Primjer funkcije $f \in \mathcal{F}$ i pripadne funkcije $f^*$ za $m = 8$.

Kako je $\|f\|_{\text{Lip}} \leq 1$, slijedi da je

$$f(x_{i+1}) \in \left[ f(x_i) - \frac{1}{m}, f(x_i) + \frac{1}{m} \right] \implies f^*(x_{i+1}) \in \left\{ f^*(x_i), f^*(x_i) - \frac{1}{m}, f^*(x_i) + \frac{1}{m} \right\},$$

pa je $\|f^*\|_{\text{Lip}} \leq 1 \implies f^* \in \mathcal{F}$. Uzmimo $x \in [0,1]$ proizvoljan. Tada za svaki $i$ vrijedi

$$|f(x) - f^*(x_i)| \leq |f(x) - f(x_i)| + |f(x_i) - f^*(x_i)| \leq \frac{1}{2m} + |x - x_i|. \tag{5.15}$$

Neka je $x_k \leq x \leq x_{k+1}$. Da bismo pokazali $\|f - f^*\|_{L^\infty} \leq \varepsilon$, razlikujemo tri slučaja.

(i) $f^*(x_{k+1}) = f^*(x_k) \implies |f(x) - f^*(x)| = \min\{|f(x) - f^*(x_k)|, |f(x) - f^*(x_{k+1})|\}$
$$\overset{(5.15)}{\leq} \tfrac{1}{2m} + \min\{|x - x_k|, |x - x_{k+1}|\} \leq \tfrac{1}{m} \leq \varepsilon.$$

(ii) $f^*(x_{k+1}) = f^*(x_k) + \frac{1}{m} \implies f(x) \overset{(5.15)}{\in} \left[ f^*(x_k + 1) - \tfrac{1}{2m} + x_{k+1} - x, f^*(x_k) + \tfrac{1}{2m} + x - x_k \right]$
$$= \left[ f^*(x) - \tfrac{1}{2m}, f^*(x) + \tfrac{1}{2m} \right]$$
$$\implies |f(x) - f^*(x)| \leq \varepsilon.$$

(iii) $f^*(x_{i+1}) = f^*(x_i) - \frac{1}{m}$, slično kao (ii).

Dakle, $\|f - f^*\|_{L^\infty} \leq \varepsilon$, pa je $\{f^* : f \in \mathcal{F}\}$ $\varepsilon$ - mreža za $\mathcal{F}$. Zato imamo gornju ogradu

$$\mathcal{N}(\mathcal{F}, \|\cdot\|_{L^\infty}, \varepsilon) \leq |\{f^* : f \in \mathcal{F}\}| \leq (m+1) \cdot 3^m \leq \left(\left\lceil \frac{1}{\varepsilon} \right\rceil + 1\right) 3^{\lceil \frac{1}{\varepsilon} \rceil} \leq \left(\frac{1}{\varepsilon} + 2\right) 3^{\frac{1}{\varepsilon}+1}.$$

Lako se pokaže da je funkcija $x \mapsto 3^{1/x} - \left(\frac{1}{x} + 2\right)$ padajuća na $\langle 0, 1]$ i ima nultočku u $x = 1$. Zato za $x \in \langle 0, 1]$ imamo $3^{1/x} \geq \frac{1}{x} + 2$. Konačno dobivamo

$$\mathcal{N}(\mathcal{F}, \|\cdot\|_{L^\infty}, \varepsilon) \leq \left(\frac{1}{\varepsilon} + 2\right) 3^{\frac{1}{\varepsilon}+1} \leq 3^{\frac{2}{\varepsilon}+1} = 3 \cdot e^{\frac{2\ln(3)}{\varepsilon}}. \qquad \square$$

U nastavku ćemo se baviti empirijskim procesima indeksiranim familijom klasifikatora $\mathcal{F} \subseteq \{0, 1\}^{\mathcal{X}}$. Sljedeće dvije tehničke tvrdnje trebat će nam u dokazima u nastavku.

**Napomena 5.17.** *Neka su* $X, Y$ *nezavisne slučajne varijable. Neka su* $f : \mathbb{R}^2 \to \mathbb{R}$ *i* $g : \mathbb{R} \to \mathbb{R}$ *Borelove i neka je* $\mathbb{E}[|f(X, Y)|] < \infty$ *i* $\mathbb{E}[|g(X)|] < \infty$. *Da bismo pokazali nejednakost*

$$\mathbb{E}[f(X, Y)] \leq \mathbb{E}[g(X)],$$

*dovoljno ju je pokazati za fiksan X, odnosno da vrijedi*

$$\mathbb{E}[f(x, Y)] \leq g(x) \qquad \forall x \in \mathbb{R}.$$

*Ako to pokažemo, koristeći nezavisnost i Fubinijev teorem, slijedit će*

$$\begin{aligned}\mathbb{E}[f(X, Y)] &= \int_{\mathbb{R}^2} f(x, y) d\mathbb{P}_{X,Y}(x, y) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} f(x, y) d\mathbb{P}_Y(y) \right) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} \mathbb{E}[f(x, Y)] d\mathbb{P}_X(x) \leq \int_{\mathbb{R}} g(x) d\mathbb{P}_X(x) = \mathbb{E}[g(X)].\end{aligned}$$

*Analogna tvrdnja vrijedi za obratnu nejednakost, kao i za slučajne vektore* $X, Y$.

**Lema 5.18.** *Neka su X, Y nezavisni slučajni vektori takvi da je* $\mathbb{E}[Y] = 0$ *i neka je* $f : \mathbb{R}^n \to \mathbb{R}$ *izmjeriva konveksna funkcija. Ako su* $f(X)$ *i* $f(X + Y)$ *integrabilne, vrijedi*

$$\mathbb{E}[f(X)] \leq \mathbb{E}[f(X + Y)].$$

*Dokaz.* Neka je $x \in \mathbb{R}^n$ proizvoljan. Tada je i funkcija $y \mapsto f(x+y)$ konveksna, pa koristeći Jensenovu nejednakost dobivamo

$$f(x) = f(x + \mathbb{E}[Y]) = f(\mathbb{E}[x + Y]) \leq \mathbb{E}[f(x + Y)].$$

Nejednakost sada slijedi po Napomeni 5.17. $\square$

Za familiju klasifikatora $\mathcal{F}$, u metričkom prostoru $\left(\mathcal{F}, \|\cdot\|_{L^2(\mu)}\right)$ imamo sljedeću ogradu na brojeve natkrivanja koja je eksponencijalna u $\mathrm{vc}(\mathcal{F})$ i ne ovisi o $\mu$.

**Teorem 5.19.** *Neka je $(\Omega, \Sigma, \mu)$ vjerojatnosni prostor i $\mathcal{F} \subseteq \{0,1\}^{\Omega}$. Tada za svaki $\varepsilon \in \langle 0, 1\rangle$ vrijedi*

$$\mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(\mu)}, \varepsilon) \leq \left(\frac{2}{\varepsilon}\right)^{C\cdot \mathrm{vc}(\mathcal{F})},$$

*gdje je $C > 0$ apsolutna konstanta.*

*Dokaz.* Vidi [7, Theorem 8.3.18]. □

Sljedeća lema daje ogradu za supremum epirijskog procesa preko Rademacherove kompleksnosti familije $\mathcal{F}$. Slijedi da ako $\mathcal{F}$ ima malu Rademacherovu kompleksnost, $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$ će dobro aproksimirati $\mathbb{E}[f(X)]$ uniformno po svim $f$.

**Lema 5.20** (Simetrizacija za empirijske procese)**.** *Neka su $X_1, \dots, X_n$ nezavisne i jednako distribuirane slučajne varijable i neka je $\mathcal{F}$ familija Borelovih funkcija. Tada vrijedi*

$$\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] \leq 2\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} \sigma_i f(X_i)\right|\right],$$

*gdje su $\sigma_1, \dots, \sigma_n$ nezavisne Rademacherove slučajne varijable, nezavisne od $X_1, \dots, X_n$.*

*Dokaz.* Pretpostavimo da je $\mathcal{F}$ konačna. Neka su $X'_1, \dots, X'_n$ nezavisne kopije od $X_1, \dots, X_n$. Označimo

$$Z_{f,i} := \frac{1}{n}(f(X_i) - \mathbb{E}[f(X)]) \quad \text{i} \quad Z'_{f,i} := \frac{1}{n}(f(X'_i) - \mathbb{E}[f(X)]).$$

Izraz $\max_{f\in\mathcal{F}}\left|\sum_{i=1}^{n} Z_{f,i}\right|$ je konveksna funkcija slučajnog vektora $(Z_{f,i})_{f\in\mathcal{F},1\leq i\leq n}$, a slučajni vektori $(Z_{f,i})_{f\in\mathcal{F},1\leq i\leq n}$ i $(-Z'_{f,i})_{f\in\mathcal{F},1\leq i\leq n}$ su nezavisni s očekivanjem 0, pa primjenom Leme 5.18 dobivamo

$$\begin{aligned}\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] &= \mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\sum_{i=1}^{n} Z_{f,i}\right|\right] \overset{5.18}{\leq} \mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\left(Z_{f,i} - Z'_{f,i}\right)\right|\right] \\ &= \mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - f(X'_i))\right|\right].\end{aligned}$$

Primijetimo da su slučajni vektori $(f(X_i))_{f\in\mathcal{F}}$ i $(f(X'_i))_{f\in\mathcal{F}}$ nezavisni i jednako distribuirani, za svaki $i$.

$\Longrightarrow$ Vektor $(f(X_i) - f(X'_i))_{f\in\mathcal{F}}$ je simetričan, za svaki $i$.

$\Longrightarrow$ Vektori $(f(X_i) - f(X'_i))_{f\in\mathcal{F}}$ i $(\sigma_i(f(X_i) - f(X'_i)))_{f\in\mathcal{F}}$ su jednako distribuirani, za svaki $i$.

$\overset{\text{nezavisnost}}{\Longrightarrow}$ Vektori $(f(X_i) - f(X'_i))_{f\in\mathcal{F},1\leq i\leq n}$ i $(\sigma_i(f(X_i) - f(X'_i)))_{f\in\mathcal{F},1\leq i\leq n}$ su jednako distribuirani.

Sada imamo

$$\begin{aligned}\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}(f(X_i)-f(X_i'))\right|\right] &= \mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i(f(X_i)-f(X_i'))\right|\right]\\ &\leq \mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i)\right|\right]+\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i')\right|\right]\\ &= 2\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i)\right|\right].\end{aligned}$$

Kako smo nejednakost pokazali za konačne familije $\mathcal{F}$, za beskonačne familije nejednakost slijedi direktno računanjem supremuma po svim konačnim podfamilijama. □

Sljedeći teorem daje ogradu za Rademacherovu kompleksnost preko VC dimenzije familije $\mathcal{F}$. Posebno, govori da uniformni zakon velikih brojeva iz Teorema 5.16 također vrijedi za sve familije klasifikatora $\mathcal{F}$ s konačnom VC dimenzijom.

**Teorem 5.21.** *Neka su $X, X_1, \dots, X_n$ nezavisne i jednako distribuirane slučajne varijable. Neka je $\mathcal{F}$ familija Borelovih funkcija s vrijednostima u $\{0,1\}$ i konačnom VC dimenzijom $\mathrm{vc}(\mathcal{F}) \geq 1$. Tada vrijedi*

$$\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|\right] \leq 2\mathbf{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(X_i)\right|\right] \lesssim \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}.$$

*gdje su $\sigma_1, \dots, \sigma_n$ nezavisne Rademacherove slučajne varijable, nezavisne od $X_1, \dots, X_n$.*

Pokažimo prvo jednu jednostavnu lemu koja će nam trebati za dokaz.

**Lema 5.22.** *Neka je $\mathcal{X}$ proizvoljan skup, neka je $\mathcal{F} \subseteq \{0,1\}^{\mathcal{X}}$ proizvoljna familija funkcija i $f \in \{0,1\}^{\mathcal{X}}$. Tada vrijedi*

$$\mathrm{vc}(\mathcal{F} \cup \{f\}) \leq \mathrm{vc}(\mathcal{F}) + 1.$$

*Dokaz.* Ako je $f \in \mathcal{F}$ ili $\mathrm{vc}(\mathcal{F}) = \infty$, tvrdnja je očita. Pretpostavimo $f \notin \mathcal{F}$ i $\mathrm{vc}(\mathcal{F}) < \infty$. Označimo $\mathcal{G} := \mathcal{F} \cup \{f\}$ i pretpostavimo suprotno: $\mathrm{vc}(\mathcal{G}) \geq \mathrm{vc}(\mathcal{F}) + 2$. Neka je $A \subseteq \mathcal{X}$, $|A| = \mathrm{vc}(\mathcal{F}) + 2$ neki skup koji se može rastaviti sa $\mathcal{G}$. Neka je $x \in A$ proizvoljan i neka je $B = A \setminus \{x\}$. Tada za svaku $g \in \{0,1\}^B$ postoji $h \in \mathcal{G}$ takva da je $h|_G = g$ i $h(x) \neq f(x)$. Zato se $B$ može rastaviti sa $\mathcal{F}$, pa je $\mathrm{vc}(\mathcal{F}) \geq |B| = |A| - 1 = \mathrm{vc}(\mathcal{F}) + 1$. Kontradikcija! □

*Dokaz Teorema 5.21.* Prvu nejednakost pokazali smo u Lemi 5.20. Pokazujemo drugu nejednakost. Primijetimo da za $\mathcal{G} \subseteq \mathcal{F}$, $|\mathcal{G}| < \infty$ imamo

$$\sqrt{\frac{\mathrm{vc}(\mathcal{G})}{n}} \leq \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}.$$

Zato, po Napomeni 5.2, možemo bez smanjenja općenitosti pretpostaviti da je familija $\mathcal{F}$ konačna. Možemo pretpostaviti i $0 \in \mathcal{F}$, jer ako pokažemo teorem za familiju $\mathcal{F} \cup \{0\}$, vrijedit će

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] \leq \mathbb{E}\left[\max_{f\in\mathcal{F}\cup\{0\}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right|\right] \lesssim \sqrt{\frac{\mathrm{vc}(\mathcal{F}\cup\{0\})}{n}}$$

$$\overset{5.22}{\leq} \sqrt{\frac{\mathrm{vc}(\mathcal{F})+1}{n}} \leq \sqrt{\frac{2\,\mathrm{vc}(\mathcal{F})}{n}} \lesssim \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}.$$

Označimo

$$Z_f = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i f(X_i).$$

Tada je nejednakost koju pokazujemo ekvivalentna

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}|Z_f|\right] \lesssim \sqrt{\mathrm{vc}(\mathcal{F})}.$$

Prema Napomeni 5.17, dovoljno je pokazati da gornja nejednakost vrijedi za "fiksne" $X_1, \ldots, X_n$. Neka su $x_1, \ldots, x_n \in \mathbb{R}$ proizvoljni i fiksni. Označimo

$$\hat{Z}_f := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i f(x_i).$$

Tada za proizvoljne $f, g \in \mathcal{F}$ vrijedi

$$\begin{aligned}\left\|\hat{Z}_f - \hat{Z}_g\right\|_{\psi_2}^2 &= \frac{1}{n}\left\|\sum_{i=1}^{n}\sigma_i(f-g)(x_i)\right\|_{\psi_2}^2 \overset{2.17}{\lesssim} \frac{1}{n}\sum_{i=1}^{n}\|\sigma_i(f-g)(x_i)\|_{\psi_2}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(f-g)(x_i)^2\|\sigma_i\|_{\psi_2}^2 \overset{2.16(a)}{\lesssim} \frac{1}{n}\sum_{i=1}^{n}(f-g)(x_i)^2 \\ &= \|f-g\|_{L^2(\mu_n)}^2,\end{aligned}$$

gdje smo sa $\mu_n$ označili vjerojatnosnu mjeru na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definiranu sa

$$\mu_n(A) = \frac{|\{i : x_i \in A\}|}{n}.$$

Primijetimo da vrijedi

$$\|f-g\|_{L^2(\mu_n)}^2 = \frac{1}{n}\sum_{i=1}^{n}\underbrace{(f(x_i)-g(x_i))^2}_{\in\{0,1\}} \leq 1.$$

Dakle, u metričkom prostoru $(\mathcal{F}, \|\cdot\|_{L^2(\mu_n)})$ imamo $\operatorname{diam}(\mathcal{F}) \le 1$ i $\left\|\hat{Z}_f - \hat{Z}_g\right\|_{\psi_2} \lesssim \|f - g\|_{L^2(\mu_n)}$. Sada smo u uvjetima Dudleyevog teorema i Napomene 5.11, pa imamo

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\hat{Z}_f\right|\right] = \mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\hat{Z}_f - \hat{Z}_0\right|\right] \overset{5.11}{\lesssim} \int_0^1 \sqrt{\log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(\mu_n)}, \varepsilon)}d\lambda(\varepsilon). \tag{5.16}$$

Iz Teorema 5.19 slijedi

$$\log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(\mu_n)}, \varepsilon) \lesssim \operatorname{vc}(\mathcal{F}) \ln\left(\frac{2}{\varepsilon}\right), \quad \text{za svaki } \varepsilon \in \langle 0, 1\rangle.$$

Konačno, spajanjem dobivenih nejednakosti dobivamo

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\hat{Z}_f\right|\right] \lesssim \sqrt{\operatorname{vc}(\mathcal{F})}\underbrace{\int_0^1 \sqrt{\ln\left(\frac{2}{\varepsilon}\right)}d\lambda(\varepsilon)}_{<\infty} \lesssim \sqrt{\operatorname{vc}\mathcal{F}}. \qquad \square$$

Kod aproksimacije funkcije distribucije empirijskom funkcijom distribucije, važan je Glivenko-Cantellijev teorem, koji kaže da empirijske funkcije distribucije gotovo sigurno uniformno konvergiraju prema teoretskoj funkciji distribucije. Sljedeći teorem govori o grešci u toj uniformnoj aproksimaciji.

**Teorem 5.23.** *Neka su $X, X_1, \dots, X_n$ nezavisne slučajne varijable s funkcijom distribucije $F$. Označimo sa $F_n$ empirijsku funkciju distribucije*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\langle -\infty, x]}(X_i).$$

*Tada vrijedi*

$$\mathbf{E}\left[\sup_{x\in\mathbb{R}}|F_n(x) - F(x)|\right] \lesssim \frac{1}{\sqrt{n}}.$$

*Dokaz.* Ovo je direktna posljedica Teorema 5.21. Neka je $\mathcal{F} = \{\mathbb{1}_{\langle -\infty, x]} : x \in \mathbb{R}\}$. Prema primjeru 1.5 vrijedi $\operatorname{vc}(\mathcal{F}) \le 2$. Koristeći $F(x) = \mathbb{E}[\mathbb{1}_{\langle -\infty, x]}(X)]$, imamo

$$\mathbf{E}\left[\sup_{x\in\mathbb{R}}|F_n(x) - F(x)|\right] = \mathbf{E}\left[\sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\langle -\infty, x]}(X_i) - \mathbb{E}[\mathbb{1}_{\langle -\infty, x]}(X)]\right|\right] \overset{5.21}{\lesssim} \sqrt{\frac{\operatorname{vc}(\mathcal{F})}{n}} \lesssim \frac{1}{\sqrt{n}}. \quad \square$$

Prisjetimo se definicija iz 1. poglavlja. Funkciju koja minimizira rizik $R$ označili smo sa $f^*$, a slučajnu funkciju koja minimizira empirijski rizik $R_n$ sa $f_n^*$. Sljedeći teorem daje ogradu na razliku u grešci optimalne funkcije veze

$$R(f_n^*) - R(f^*)$$

koja je nastala zbog učenja na konačnom uzorku. Također, govori i o točnosti aproksimacije rizika empirijskim rizikom.

**Teorem 5.24.** *Neka je $Y = \{0, 1\}$, neka je $l$ funkcija 0-1 gubitka i neka $\mathcal{F}$ ima konačnu VC dimenziju* $\mathrm{vc}(\mathcal{F}) \geq 1$. *Tada vrijedi*

$$\mathbb{E}[R(f_n^*)] - R(f^*) \leq 2\mathbf{E}\left[\sup_{f\in\mathcal{F}}|R_n(f) - R(f)|\right] \lesssim \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}.$$

Za dokaz teorema će nam trebati sljedeća lema.

**Lema 5.25.** *Neka je $\mathcal{X}$ proizvoljan skup, $Y = \{0, 1\}$ i neka je $l$ funkcija 0-1 gubitka. Neka je $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$ i neka je familija $\mathcal{G} \subseteq \{0, 1\}^{X\times Y}$ dana sa*

$$\mathcal{G} = \{(x, y) \mapsto l(f(x), y) : f \in \mathcal{F}\}. \tag{5.17}$$

*Tada vrijedi* $\mathrm{vc}(\mathcal{F}) = \mathrm{vc}(\mathcal{G})$.

*Dokaz.* Neka je $A = \{(x_1, y_1), \ldots, (x_n, y_n)\} \in \mathcal{X} \times \mathcal{Y}$ proizvoljni. Vrijede ekvivalencije

$$\begin{aligned}
&(x_i, y_i) \text{ su međusobno različiti i } \{(x_1, y_1), \ldots, (x_n, y_n)\} \text{ se može rastaviti sa } \mathcal{G}\\
&\iff \{(l(f(x_i), y_i))_{i=1}^n : f \in \mathcal{F}\} = \{0, 1\}^n\\
&\iff \{(\neg(f(x_i) \oplus y_i)))_{i=1}^n : f \in \mathcal{F}\} = \{0, 1\}^n\\
&\iff \{(f(x_i) \oplus y_i)_{i=1}^n : f \in \mathcal{F}\} = \{0, 1\}^n\\
&\iff \{((f(x_i) \oplus y_i) \oplus y_i)_{i=1}^n : f \in \mathcal{F}\} = \{0, 1\}^n\\
&\iff \{(f(x_i))_{i=1}^n : f \in \mathcal{F}\} = \{0, 1\}^n\\
&\iff x_i \text{ su međusobno različiti i } \{x_1, \ldots, x_n\} \text{ se može rastaviti sa } \mathcal{F},
\end{aligned}$$

gdje smo na nule i jedinice primijenili logičke operatore $\neg$ (negacija) i $\oplus$ (isključivo ili). Sada iz definicije VC dimenzije slijedi $\mathrm{vc}(\mathcal{F}) = \mathrm{vc}(\mathcal{G})$. □

*Dokaz Teorema 5.24.* Neka je $\mathcal{G} \subseteq \mathcal{F}$, $|\mathcal{G}| < \infty$ proizvoljna. Tada po točkama vrijedi

$$\begin{aligned}
R\left(\arg\min_{f\in\mathcal{G}} R_n(f)\right) &\leq R_n\left(\arg\min_{f\in\mathcal{G}} R_n(f)\right) + \max_{f\in\mathcal{G}\cup\{f^*\}}|R_n(f) - R(f)|\\
&\leq R_n(f^*) + \max_{f\in\mathcal{G}\cup\{f^*\}}|R_n(f) - R(f)|\\
&\leq R(f^*) + 2\max_{f\in\mathcal{G}\cup\{f^*\}}|R_n(f) - R(f)|.
\end{aligned}$$

Zato je

$$
\begin{aligned}
\mathbb{E}[R(f_n^*)] = \mathbb{E}\left[R\left(\operatorname*{arg\,min}_{f\in\{f_n^*\}} R_n(f)\right)\right] &\leq \sup_{\substack{\mathcal{G}\subseteq\mathcal{F}\\|\mathcal{G}|<\infty}} \mathbb{E}\left[R\left(\operatorname*{arg\,min}_{f\in\mathcal{G}} R_n(f)\right)\right] \\
&\leq R(f^*) + 2\sup_{\substack{\mathcal{G}\subseteq\mathcal{F}\\|\mathcal{G}|<\infty}} \mathbb{E}\left[\max_{f\in\mathcal{G}\cup\{f^*\}} |R_n(f) - R(f)|\right] \\
&\leq R(f^*) + 2\sup_{\substack{\mathcal{G}\subseteq\mathcal{F}\\|\mathcal{G}|<\infty}} \mathbb{E}\left[\max_{f\in\mathcal{G}} |R_n(f) - R(f)|\right],
\end{aligned}
$$

pa je prva nejednakost pokazana. Pokažimo drugu nejednakost. Primijetimo da vrijedi

$$
\begin{aligned}
\mathbf{E}\left[\sup_{f\in\mathcal{F}} |R_n(f) - R(f)|\right] &= \mathbf{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n l(f(X_i), Y_i) - \mathbb{E}[l(f(X), Y)]\right|\right] \\
&= \mathbf{E}\left[\sup_{h\in\mathcal{H}} \left|\frac{1}{n}\sum_{i=1}^n h(Z_i) - \mathbb{E}[h(Z)]\right|\right],
\end{aligned}
$$

gdje je $\mathcal{H}$ dana kao u (5.17), $Z = (X, Y)$ i $Z_i = (X_i, Y_i)$. Iz Leme 5.25 slijedi $1 \leq \mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{F}) < \infty$. Primjenom Teorema 5.21 na slučajne vektore $Z, Z_1, \ldots, Z_n$ i familiju $\mathcal{H}$ dobivamo

$$
\mathbf{E}\left[\sup_{h\in\mathcal{H}} \left|\frac{1}{n}\sum_{i=1}^n h(Z_i) - \mathbb{E}[h(Z)]\right|\right] \lesssim \sqrt{\frac{\mathrm{vc}(\mathcal{H})}{n}} = \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}. \qquad \square
$$

Prethodni teorem nam ujedno govori koliko velik uzorak trebamo uzeti kako bi rizik $R(f_n^*)$ bio blizu optimalnog rizika $R(f^*)$. Ako želimo da vrijedi

$$
\mathbb{E}[R(f_n^*)] - R(f^*) \leq \varepsilon,
$$

dovoljno je uzeti veličinu uzorka $n$ proporcionalnu $\frac{\mathrm{vc}(\mathcal{F})}{\varepsilon^2}$. Ako želimo povećati kompleksnost familije $\mathrm{vc}(\mathcal{F})$, očekujemo da ćemo trebati proporcionalno povećati i uzorak.

U svim rezultatima do sada trebala nam je konačna VC dimenzija. Sljedeći primjer pokazuje da je konačnost VC dimenzije karakterizacija uniformnih Glivenko-Cantellijevih klasa.

**Primjer 5.26** (Karakterizacija uniformne Glivenko-Cantellijeve klase)**.** *Neka je $\mathcal{F}$ familija Borelovih funkcija s vrijednostima u $\{0, 1\}$. Ako $\mathcal{F}$ ima konačnu VC dimenziju, uz određene tehničke pretpostavke za izmjerivost, iz Teorema 5.21 i Markovljeve nejednakosti slijedi*

$$
\mathbb{P}\left(\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right| > \varepsilon\right) \lesssim \frac{\sqrt{\mathrm{vc}(\mathcal{F})}}{\varepsilon\sqrt{n}}.
$$

*Kako gornja ograda ne ovisi o distribuciji $\mu$ od X, imamo uniformnu konvergenciju po $\mu$*

$$\lim_{n\to\infty} \sup_{\mu} \mathbb{P}\left( \sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)] \right| > \varepsilon \right) = 0, \tag{5.18}$$

*odnosno $\mathcal{F}$ je uniformna Glivenko-Cantellijeva klasa.*
*Obrat lako možemo pokazati konstrukcijom. Pretpostavimo da $\mathcal{F}$ ima beskonačnu VC dimenziju. Neka je $n \in \mathbb{N}$ proizvoljan i neka je $\{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ neki skup koji se može rastaviti sa $\mathcal{F}$. Neka je $f_{\mathbf{X}} \in \mathcal{F}$ slučajna funkcija takva da za $i = 1, \ldots, n$ vrijedi*

$$f_{\mathbf{X}}(x_i) = \mathbb{1}_{\{X_1,\ldots,X_n\}}(x_i).$$

*Ako $X_1, \ldots, X_n$ imaju uniformnu distribuciju na $\{x_1, \ldots, x_n\}$, slijedi*

$$\begin{aligned}
\underbrace{\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)] \right|\right]}_{=:Z} &\geq \mathbb{E}\left[\left| \frac{1}{n} \sum_{i=1}^{n} f_{\mathbf{X}}(X_i) - \mathbb{E}[f_{\mathbf{X}}(X)] \right|\right] = \mathbb{E}\left[1 - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_1,\ldots,X_n\}}(x_i)\right] \\
&= 1 - \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(x_i \in \{X_1, \ldots, X_n\}) = 1 - \mathbb{P}(x_1 \in \{X_1, \ldots, X_n\}) \\
&= 1 - \left(1 - \frac{1}{n}\right)^n \geq 1 - \frac{1}{e} \geq \frac{1}{2}.
\end{aligned}$$

*Sada lako dobivamo*

$$\frac{1}{2} \leq \mathbb{E}[Z] = \mathbb{E}\Big[Z \mathbb{1}_{\{Z\leq\frac{1}{4}\}}\Big] + \mathbb{E}\Big[Z \mathbb{1}_{\{Z>\frac{1}{4}\}}\Big] \leq \frac{1}{4}\mathbb{P}\Big(Z \leq \frac{1}{4}\Big) + \mathbb{P}\Big(Z > \frac{1}{4}\Big) = \frac{1}{4} + \frac{3}{4}\mathbb{P}\Big(Z > \frac{1}{4}\Big),$$

*iz čega slijedi da za uniformnu distribuciju na $\{x_1, \ldots, x_n\}$ vrijedi $\mathbb{P}\big(Z > \frac{1}{4}\big) \geq \frac{1}{3}$, pa limes* (5.18) *za $\varepsilon = \frac{1}{4}$ ne može biti* 0.

# Bibliografija

[1] Stéphane Boucheron, Gábor Lugosi i Pascal Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford university press, 2013.

[2] Luc Devroye i László Györfi, *Nonparametric Density Estimation: The L1 View*, Wiley, 1985.

[3] Charles M. Goldie i Claudia Klüppelberg, *Subexponential distributions*, A practical guide to heavy tails: statistical techniques and applications (1998), 435–459.

[4] Gábor Lugosi, *Concentration-of-measure inequalities*, `http://www.econ.upf.edu/~lugosi/anu.pdf`, 2009.

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

[6] Vladimir Vapnik, *The nature of statistical learning theory*, Springer science & business media, 1999.

[7] Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, sv. 47, Cambridge university press, 2018.

[8] Martin J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, sv. 48, Cambridge University Press, 2019.

# Sažetak

U ovom radu prikazali smo neke metode dokazivanja koncentracijskih nejednakosti i njihove primjene u teoriji statističkog učenja.

Najjednostavniji i najviše proučavani primjer su sume nezavisnih slučajnih varijabli. Usporedili smo što o njihovim fluktuacijama govore Centralni granični teorem i neasimptotske koncentracijske nejednakosti. Chernoffljevom metodom pokazali smo Hoeffdingovu i Bernsteinovu nejednakost za sume ograničnih varijabli. Zatim smo definirali subgaussovske i subeksponencijalne varijable te pokazali da za njih također vrijede slične nejednakosti.

Općenitiji primjer su generalne funkcije nezavisnih slučajnih varijabli. Koristeći martingalnu metodu pokazali smo Efron-Steinovu i McDiarmidovu nejednakost. Rezultate smo primijenili na funkcije sa svojstvom omeđenih razlika i samoomeđujućim svojstvom, kao što su minimum empirijskog rizika i Rademacherova usrednjenja. Uveli smo osnovne pojmove iz teorije informacija i dokazali subaditivnost $\Phi$-entropije, iz koje slijede Logaritamske Sobolevljeve nejednakosti. Zatim smo pokazali kako se iz njih metodom entropije mogu dobiti eksponencijalne koncentracijske nejednakosti.

U raznim primjenama važno je razumijevanje fluktuacija supremuma slučajnih procesa, a posebno empirijskih procesa. Metodom ulančavanja pokazali smo Dudleyevu nejednakost, koja daje ogradu na očekivanje supremuma preko metričke entropije. Iz nje smo dobili ogradu na očekivanje supremuma empirijskog procesa preko VC dimenzije familije indeksa. Neke od posljedica tih rezultata su uniformni zakon velikih brojeva za Lipschitzove funkcije i ograda na razliku rizika i empirijskog rizika.

# Summary

In this thesis, we present some methods for proving concentration inequalities and their applications in the theory of statistical learning.

The simplest and most studied examples are sums of independent random variables. We compared what the Central Limit Theorem and non-asymptotic concentration inequalities say about their fluctuations. Using Chernoff bounds, we have shown that Hoeffding and Bernstein inequalities hold for sums of bounded variables. We then defined subgaussian and subexponential random variables and showed that similar inequalities hold for such variables.

A more general example is that of general functions of independent random variables. Using the martingale method, we showed Efron-Stein and McDiarmid inequality. We applied the results to functions with bounded differences and self-bounding functions, such as the minimum of the empirical loss and Rademacher averages. We have introduced basic concepts from information theory and proved the subadditivity of $\Phi$-entropy, from which the Logarithmic Sobolev inequalities follow. Then we showed how exponential concentration inequalities can be obtained from them by the entropy method.

In various applications, it is important to understand the fluctuations of the supremum of random processes, especially empirical processes. Using the chaining method we have shown the Dudley inequality, which gives a bound on the expectation of supremum using metric entropy. We used the inequality to show a bound for the expected supremum of an empirical process via VC dimension of the index family. Some of the consequences of these results are a uniform law of large numbers for Lipschitz functions and a bound on the difference between the loss and empirical loss.

# Životopis

Rođen sam 14. ožujka 1998. godine u Zagrebu. Pohađao sam Osnovnu školu Granešina, a 2012. sam upisao XV. gimnaziju u Zagrebu. Tijekom osnovne i srednje škole sudjelovao sam na državnim natjecanjima iz matematike i logike.

Na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu 2016. godine upisujem preddiplomski sveučilišni studij Matematika, a 2019. godine diplomski studij Matematička statistika. Tijekom studija sudjelovalo sam u radu udruge Mladi nadareni matematičari "Marin Getaldić". Sudjelovao sam i na međunarodnim studentskim natjecanjima: osvojio sam dvije druge nagrade na natjecanju IMC i 7. mjesto na natjecanju Vojtěch Jarník. Za izniman uspjeh u studiju dodijeljene su mi dvije nagrade Matematičkog odsjeka, dvije nagrade Prirodoslovno-matematičkog fakulteta te Stipendija Grada Zagreba za izvrsnost.