1 Uvod - opis teme

Dandanes živimo v dobi podatkov. Podjetja ter ustanove zbirajo podatke o svojih uporabnikih in jih nato uporabijo za izboljšanje algoritmov oz. tehnologije. Posledično postaja vse bolj pomembno vprašanje varovanja podatkov. Kako omogočiti strokovnjakom in inženirjem dostop do velikih podatkovnih baz, hkrati pa ohraniti zasebnost posameznikov? Kako lahko podjetja uporabijo vse podatke, ki jih zbirajo o svojih uporabnikih, ne da bi s tem ogrozili njihove privatnosti? Obstaja veliko različnih pristopov (najbolj pogoste so t. i. 'metode anonimizacije'), vendar se je izkazalo, da je veliko od teh neprimernih in so izpostavljeni različnim napadom. Še posebej velja izpostaviti t. i. 'background/auxiliary' napad, pri katerem napadalec uporabi dodatno podatkovno bazo za razkritje identite posameznikov v prvotni bazi (Netflix, American health records). V ta namen se je v zadnjih 15 letih pojavil koncept diferencirane zasebnosti, ki postavlja vprašanje varovanja podatkov v okvir matematične teorije. Vnaša 'naključnost' v podatkovno bazo; odgovori na poizvedbe tako niso več deterministični, ampak so probabilstični. Osnovna ideja je, da se odgovor na poizvedbo ne bo spremenil, če podatki o enem konkretnem posamezniku so ali niso v bazi. Na ta način omogočimo dostop do globalnih lastnosti celotne populacije, zaščitimo pa konkretne informacije o posameznikih. Mnogi vidijo diferencirano zasebnost kot odlično rešitev, uporabljajo jo tudi podjetja kot so Apple in Uber. Na drugi strani pa obstaja tudi precej kritik; ena glavnih izmed njih je, da koncept diferencirane zasebnosti sicer ponuja lepe matematične in teoretične temelje, ni pa zares uporaben v praksi.

Pomembno je tudi razumevanje, da se diferencirana zasebnost ne ukvarja z zaščito podatkov samih. Npr. ne ponuja odgovora na problem, kako podatke varno shraniti na nekem serverju (da jih zaščitimo pred hekerskemi vdori ipd.). Ukvarja se s tem, kako ohraniti zasebnost pri procesu objavljanja (data publishing) in pridobivanja podatkov (data mining). Za lažje razumevanje pred nadaljevanjem navedimo še dva primera. Prvi se je pojavil leta 2000, ko so raziskovalci ugotovili, da ameriške bolnišnice niso zaščitile javno dostopnih podatkov na primeren način. Kar so naredili (bolnišnice) je, da so uporabili metodo anonimizacije, to pomeni, da so iz podatkov odstranili vse eksplicitne indikatorje (npr. ime in EMŠO). Podatki so tako vsebovali le npr. spol pacienta, rojstni datum, poštno številko (ZIP) in zdravstveno stanje (poenostavljen primer). Izkazalo se je, da lahko te podatke združimo (angl. cross-reference) s podatki iz volilnega sistema, in na ta način razkrijemo identiteto pacientov. En možen način kako se lahko reši tak problem bi bil, da bolnišnice sploh ne bi javno objavile baze podatkov, ampak bi dostop do baze raziskovalcem omogočile le prek dovoljenih poizvedb. Npr. znanstveniki bi lahko 'vprašali' koliko odstotkov oseb ženskega spola ima to in to bolezen, ne bi pa smeli povprašati po diagnozi konkretnega posameznika. Pri takem pristopu se naravno pojavi vprašanje, točno katere poizvedbe dovolimo. Po eni strani izbire nočemo preveč omejiti (znanstvenikom želimo omogočiti kvalitetne podatke), po drugi strani ne želimo izpostaviti nobenih konkretnih informacij o posamezniku. Hkrati se

Pacient	**Diabetes**
Anja	1
Bojan	1
Cene	0
Darja	0
Edi	1

Tabela 1: Podatkovna baza z imeni pacientov in podatki o diabetesu.

tudi izkaže, da lahko v primeru deterministič nih poizvedb napadalec samo na podlagi odgovorov na poizvedbe sklepa o lastnostnih posameznika. Ponazorimo to s preprostim zgledom na podatkih iz tabele 1. Recimo, da bi napadalec rad izvedel Edijevo diagnozo. Privzemimo, da ima na voljo poizvedbo Q_i, ki kot odgovor vrne vsoto prvih i vrstic. Dodatno tudi ve, da se Edi nahaja na 5. mestu v tabeli. Tako bi lahko izvedel poizvedbi Q_4 in Q_5 in iz razlike obeh sklepal, da Edi ima diabetes. Odgovor na to problematiko ponuja diferencirana zasebnost, pri kateri poizvedbe postanje probabilistične in napadalec kljub dodatnim informacijam ne more priti do gotovih sklepov o Edijevi diagnozi.

Drugi primer, ki je pogost dandanes, pa se pojavi pri procesu rudarjenja podatkov (angl. data mining). Podjetja za izboljšanje svoje programske opreme, beležijo skoraj vsako našo potezo na pametnem telefonu ali računalniku. Rezultat so potem npr. sistemi za priporočanje in 'auto-correct' sistem pri tipkanju; torej algoritmi, ki za svoje delovanje potrebujejo velike količine podatkov (da se iz njih "učijo"). Podjetja zajemanje podatkov opravičujejo z izboljšanjem tehnologije. Uporabnikom je tako rečeno, da morajo žrtvovati del zasebnosti za izboljšanje tehnologije, ki jo uporabljajo. Diferencirana zasebnost ponuja možnost, da temu ni tako (v primeru, da se izkaže, da je 'skalabilna'). Tak pristop npr. že uporablja Apple, ki nekatere podatke še preden jih iz naprave (npr. iz iphona) pošlje na centralni server, 'zamaskira' s pomočjo algoritmom, ki slonijo na diferencirani zasebnosti (to so uporabili pri sistemu za predloge 'emojiev' in sistemu za pomoč pri tipkanju).

V prvem delu diplomske naloge se bom posvetil teoretičnemu ozadju. V praksi imamo opravka z različnimi vrstami podatkov, v literaturi pa se diferencirana zasebnost (kot matematični koncept) pogosto definira le za posamezno vrsto podatkov (npr. številske). Predstavil bom model, ki omogoča, da se koncept diferencirane zasebnosti definira v splošnem, tj. za vse vrste podatkov hkrati. V nadaljevanju bo podanih tudi nekaj osnovnih rezultatov, ki izhajajo iz tega modela. V drugem delu bom...

2 Matematična priprava - definicije, izreki

Naj bo Ω neprazna množica s pripadajočo algebro $\mathcal{S}$. Označimo s $\sigma(\mathcal{S})$ najmanjšo σ-algebro, ki vsebuje $\mathcal{S}$ (rečemo da S generira $\sigma(\mathcal{S})$).

Definicija (Monoton razred) Monoton razred $\mathcal{M}$ je družina podmnožic Ω (torej $\mathcal{M} \subset \mathcal{P}(\mathcal{M})$) z naslednjima lastnostima:

- $\{A_i\}_{i=1,\dots\infty} \in \mathcal{M}, A_i \subseteq A_{i+1} \Rightarrow \bigcup_{i=1,\dots\infty} A_i \in \mathcal{M}$ (zaprtost za monotono naraščajoče števne unije),
- $\{A_i\}_{i=1,\dots\infty} \in \mathcal{M}, A_i \supseteq A_{i+1} \Rightarrow \bigcap_{i=1,\dots\infty} A_i \in \mathcal{M}$ (zaprtost za monotono padajoče števne preseke).

Iz definicije takoj sledi, da je vsaka σ-algebra monoton razred (uporabimo zaprtost za poljubne števne unije in preseke). Naslednji izrek karakterizira $\sigma(\mathcal{S})$ kot najmanjši monoton razred, ki vsebuje algebro $\mathcal{S}$.

Izrek (O monotonih razredih) Naj bo $\mathcal{S}$ algebra in $\mathcal{M}$ monoton razred na množici Ω. Naj velja še $\mathcal{S} \subseteq \mathcal{M}$. Potem sledi $\sigma(\mathcal{S}) \subseteq \mathcal{M}$.

Dokaz: Označimo z $m(\mathcal{S})$ najmanjši monoton razred, ki vsebuje $\mathcal{S}$ (dobimo ga kot presek vseh monotonih razredov na Ω, ki vsebujejo $\mathcal{S}$). Ker za vsak $\mathcal{M}$ z lastnostjo $\mathcal{S} \subseteq \mathcal{M}$ velja $m(\mathcal{S}) \subseteq \mathcal{M}$, vidimo, da je dovolj pokazati $\sigma(\mathcal{S}) \subseteq m(\mathcal{S})$. Za dokaz tega je dovolj pokazati, da je $m(\mathcal{S})$ σ-algebra (sledi iz dejstva, da je $\sigma(\mathcal{S})$ najmanjša σ-algebra, ki vsebuje $\mathcal{S}$). Ker je $m(\mathcal{S})$ monoton razred, je dovolj pokazati, da je $m(\mathcal{S})$ algebra (algebra je σ-algebra natanko tedaj, ko je monoton razred; dokaz na tem mestu izpustimo).
Pokažimo najprej zaprtost za komplemente. Obravnavajmo družino množic $\mathcal{G} = \{A \mid A^c \in m(\mathcal{S})\}$. Ker je $m(\mathcal{S})$ monoton razred, sledi da je tudi $\mathcal{G}$. Dodatno velja še $\mathcal{S} \subseteq \mathcal{G}$ (sledi iz $\mathcal{S} \subseteq m(\mathcal{S})$ in dejstva, da je $\mathcal{S}$ algebra, torej zaprta za komplement). To nam zagotovi $m(\mathcal{S}) \subseteq \mathcal{G}$, s čimer smo pokazali, da je $m(\mathcal{S})$ zaprt za komplemente.
Pokažimo še zaprtost za končne unije. Definirajmo družino množic $\mathcal{H}_1 = \{A \mid A \cup B \in m(\mathcal{S}), \forall B \in \mathcal{S}\}$. Potem je $\mathcal{H}_1$ monoton razred in $\mathcal{S} \subseteq \mathcal{H}_1$. Iz minimalnosti sledi $m(\mathcal{S}) \subseteq \mathcal{H}_1$. Definirajmo še $\mathcal{H}_2 = \{B \mid A \cup B \in m(\mathcal{S}), \forall A \in m(\mathcal{S})\}$. Spet velja, da je $\mathcal{H}_2$ monoton razred. Ker je $m(\mathcal{S}) \subseteq \mathcal{H}_1$, sledi da $A \in m(\mathcal{S})$ in $B \in S$ skupaj implicirata $A \cup B \in m(\mathcal{S})$. Povedano drugače, $B \in S$ implicira $B \in \mathcal{H}_2$. Torej je $S \subseteq \mathcal{H}_2$ in iz minimalnosti dobimo $m(\mathcal{S}) \subseteq \mathcal{H}_2$, iz česar sledi, da $A, B \in m(\mathcal{S})$ implicira $A \cup B \in m(\mathcal{S})$. Torej je $m(\mathcal{S})$ res algebra in izrek je dokazan.

Na kratko ponovimo še nekaj osnovnih pojmov pri metričnih prostorih.

- Metrični prostor (D, ρ) je *končen* (angl. finite), če ima končno število elementov/točk (torej $|D| < \infty$). Primer je množica hobijev v primeru 1.

- Metrični prostor (D, ρ) je *končno-dimenzionalen* (angl. finite-dimensional), če ima končno bazo. Primer je $\mathbb{R}^n$.
- Metrični prostor (D, ρ) je *neskončno-dimenzionalen* (angl. infinite-dimensional), če nima končne baze. Primer je $C([0, 1])$.

Definicija (Kompaktnost metričnih prostorov) Metrični prostor (D, ρ) je *kompakten*, če ima vsako zaporedje v D konvergetno podzaporedje z limito prav tako v D (povedano drugače, vsako zaporedje v D ima vsaj eno stekališče vsebovano v D).

Opomba: Zgoraj podana definicija ne opisuje najbolj splošnega pojma kompaktnosti, ampak gre za ubistvu t. i. zaporedno kompaktnost (angl. sequentially compactness), kar zadostuje za potrebe tega diplomskega dela. Oba pojma kompaktnosti sta namreč ekvivalentna v primeru metričnih prostorov. Se pa pojma razlikujeta, če delamo s topološkimi prostori (kompaktnost tu definiramo drugače, prek pokritij in podpokritij).
V primeru ko je $D \subset \mathbb{R}^n$ (podmnožica Evklidskega prostora), dodatno vemo, da je D kompakten natanko tedaj, ko je zaprt in omejen.

Opomba (O kompaktnosti končnih prostorov): Naj bo (D, ρ) končen metričen prostor. Potem je D kompakten.

Dokaz opombe: Vzemimo poljubno zaporedje v D z neskočno mnogo elementi. Vidimo, da se mora vsaj en element iz D v zaporedju pojaviti neskončno mnogokrat. V nasprotnem primeru zaporedje ne bi imelo neskončno mnogo elementov (sledi iz končnosti D). Ponavljajoče se vrednosti tega elementa tvorijo podzaporedje, ki je seveda konvergentno. Torej je po zgornji definiciji D kompakten.

Definirajmo še metriko ρ_H, t.i. *Hammingovo razdaljo*, ki jo bomo potrebovali za izpeljavo nekaterih nadaljnjih rezultatov. Naj bo D poljubna množica in $a, b \in D^n$. Potem je $\rho_H(a, b)$ enaka številu mest, na katerih se vektorja razlikujeta. Opazimo, da je v 1-dimenzionalnem primeru ($n = 1$) ta metrika ekvivalentna diskretni. Morda se bralcu zastavi vprašanje, kako deluje taka metrika na vektorju npr. funkcij. Če želimo enakost skoraj povsod, lahko za D vzamemo npr. prostor L^2, torej primerjamo med seboj ekvivalenčne razrede funkcij. Metrika na prvi pogled tudi ni najbolj uporabna, saj pove le, ali sta elementa danega metričnega prostora različna in ne 'koliko' sta različna. Zanimivo je, da je to v primeru diferencirane zasebnosti primerna stvar, saj tu med sabo primerjamo podatkovne baze in ne vnosov znotraj posamezne baze.

3 Splošni podatkovni model in definicija diferencirane zasebnosti

3.1 Database model

Naj bo (U, ρ) poljuben metrični prostor in $D \subseteq U$. Posamezni vnosi v opazovani podatkovni bazi so elementi množice D. Celotno bazo prikažemo z vektorjem $\mathbf{d} = (d_1, ..., d_n) \in D^n$, kjer $d_i \in D$ predstavlja i-ti vnos oz. vrstico.

Množico U opremimo z Borelovo σ -algebro, označimo jo z $\mathcal{A}_U$, ki je najmanjša σ-algebra, ki vsebuje vse odprte množice v U. Tako σ-algebro generiramo preko metrične topologije. Za lažjo predstavo podajmo grob opis tega postopka. S pomočjo metrike ρ na U lahko definiramo odprte krogle $B_r(x) = \{y \in U \mid \rho(x, y) < r\}$. To zadošča, da lahko definiramo bazo topologije, ki je oblike $\mathcal{B} = \{B_r(x) \mid x \in U, r > 0\}$ (povedano drugače, vsak metrični prostor je hkrati topološki prostor oz. metrika nam porodi topologijo). Ko enkrat imamo topologijo (lahko si jo predstavljamo kot podmnožico potenčne množice U, ki vsebuje vse odprte množice v U), lahko le-to uporabimo za generiranje Borelove σ-algebre.

$\mathcal{A}_U$ nam potem naravno porodi $\mathcal{A}_D := \{A \in \mathcal{A}_U | A \subset D\}$ na D. Predpostavimo tudi, da je U^n (in s tem D^n) opremljen s produktno σ-algebro, ki je generirana prek $\{A_1 \times ... \times A_n | A_i \in \mathcal{A}_U\}$ in jo označimo z $\mathcal{A}_{U^n}$.

Tak model podatkovne baze je zelo splošen (U je namreč poljubni metrični prostor) in nam omogoča enotno obravnavo različnih vrst podatkov: numeričnih, kategoričnih in funkcijskih.

Primer 1 Recimo, da imamo na voljo podatkovno bazo, v kateri so zabeleženi hobiji posameznikov. Množico vseh hobijev lahko označimo s $\mathcal{H} = \{nogomet, kitara, ...\}$. Predpostavka o končnosti $\mathcal{H}$ je tu smiselna in neomejujoča. Za D potem lahko vzamemo $2^{\mathcal{H}}$. Pri izbiri metriki imamo precej proste roke, vzemimo npr. diskretno metriko $\rho(A, B) = 1$ če $A \neq B$ in 0 drugače. Borelova σ-algebra $\mathcal{A}_D$ je tu kar enaka D. Opazimo tudi, da ni nujno, da imajo vsi elementi v D^n (torej vnosi v naši podatkovni bazi) enako število elementov, kar odraža dejstvo, da nimamo vsi ljudje enakega števila hobijev.

Primer 2.a Kot primer za numerične podatke obravnavajmo barvne slike, torej $D = \mathbb{R}^{n \times m \times 3}$ (RGB slike dimenzije $n \times m$). Kot metrika se tu naravno ponuja $\rho(A, B) = \sum_{i,j,k} |a_{i,j,k} - b_{i,j,k}|$. Za Borelovo σ-algebro $\mathcal{A}_D$ vzamemo produtkno σ-algebro, torej $\mathcal{A}_D = \{A_1 \times A_2 \times A_3 | A_1 \in \mathcal{B}(\mathbb{R}^n), A_2 \in \mathcal{B}(\mathbb{R}^m), A_3 \in \mathcal{B}(\mathbb{R}^3)\}$. Ta koncept lahko razširimo na 3-D barvne slike in tudi na video posnetke.

Primer 2.b Primer mešanih podatkov nam ponuja enostavna baza zdravstvenih podatkov. Recimo, da so elementi naše baze oblike *(kvazi-identifikator pacienta,*

starost, spol, bolezen, bolezen 1, bolezen 2, ...). D potem lahko izberemo takole

$$D = \{1, 2, ..., \check{s}t.pacientov\} \times \{1, ..., 120\} \times \{M, \check{Z}\} \times \\ \{Ljubljana, ..., SpodnjiDuplek\} \times \{0, 1\}^n.$$

Kar pogosto naredimo v praksi je, da najprej kategorične podatke spremenimo v numerične (npr. z uporabo one-hot encodinga). Metrika na D in pripradajoča Borelova algebra potem izgledata podobno kot v zgornjem primeru.

Primer 3 Navedimo še primer, ko imamo opravka s t. i. funkcijskimi podatki. Te se pojavijo npr. pri merjenju porabe elektrike v gospodinjstvih, kar lahko predstavimo kot graf porabe v odvisnosti od časa. Če meritev opravimo le ob določenih časovnih točkah, lahko za D vzamemo npr. prostor zaporedij (angl. sequence space) l_∞ ali l_2. Drugače lahko vzamemo za D npr. $C([0, T])$ ali $L_2([0, T])$. Tu T označuje dolžino opazovanega časovnega obdobja. Dodajmo še, da so prostori l_p le posebni primeri prostorov L_p, ko delamo na merljivem prostoru $(\mathbb{N}, 2^{\mathbb{N}})$, za mero pa vzamemo t. i. mero štetja. Vsi ti prostori so Banachovi prostori (t.j. polni normirani metrični prostori), kar pomeni, da imajo naravno podano normo. Le-ta nam inducira metriko ρ, prav tako pa lahko prek norme pridemo bo pripadajoče Borelove σ-algebre (glej izpeljavo zgoraj prek metrične topologije).

Pravimo da sta dve podatkovni bazi, **d**= $(a_1, ..., a_n)$ in **d'** = $(b_1, ..., b_n)$, *sosednji*, če se razlikujeta v natanko enem vnosu. Torej:

- obstaja $j \in \{1, ...n\}$, da velja $a_j \neq b_j$,
- za vsak $i \in \{1, ..., n\} \setminus j$ velja $a_i = b_i$.

Sosednji bazi označimo z $\mathbf{d} \sim \mathbf{d}'$. Če obravnamo D^n kot metrični prostor s pripadajočo Hammingovo metriko ρ_H, lahko ekvivalentno rečemo, da mora veljati $\rho_H(\mathbf{d},\mathbf{d}')= 1$.

Za izpeljavo nekaterih rezultatov v nadaljevanju, moramo predpostaviti, da je D kompakten metrični prostor. V primeru kompaktnosti nato definiramo še diameter kot $diam(D) = \max_{d,d' \in D} \rho(d, d')$. Obstoj diametra je posledica dejstva, da zvezna funkcija ρ doseže svoj maksimum na kompaktni množici D.

3.2 Query model

Poizvedba (query) je način pridobitve željenih informacij iz podatkovne baze. V prejšnjem poglavju smo podatkovno bazo predstavili kot metrični prostor in enako sedaj storimo za množico vseh možnih odgovor (angl. set of all possible responses) na posamezno poizvedbo. Tak metrični prostor označimo z (E_Q, ρ_Q) in ga ponovno opremimo z Borelovo σ-algebro $\mathcal{A}_Q$ (indeks Q tu ponazarja odvisnost od poizvedbe, kar je naravno, saj različne poizvedbe vodijo do različnih

množic možnih odgovorov). Sedaj lahko definiramo poizvedbo kot merljivo funkcijo $Q : U^n \to E_Q$, torej $Q^{-1}(A) \in \mathcal{A}_{U^n}$ za vsako $A \in Q$.

Primer 4 Kot pri konstrukciji podatkovne baze $\mathbf{d} \in D^n$, imamo tudi pri izbiri prostora možnih odgovorov precej proste roke. Če se vrnemo na primer podatkovne baze hobijev, bi npr. lahko povprašali po številu ljudi, ki igrajo nogomet. Odgovor na to poizvedbo bi bil numeričen ($E_Q = \mathbb{N}$). Lahko pa bi nas zanimalo, kateri so 3 najpogostejši hobiji v bazi. Odgovor tu bi bil verjetno množica hobijev ($E_Q = 2^{\mathcal{H}}$).

Dalje lahko definiramo pojem *odzivnega mehanizma* (angl. response mechanism), ki nam omogoča, da v našo podatkovno bazo vnesemo 'naključnost' in na ta način preprečimo, da bi lahko prek poizvedb prišli do konkretnih informacij o posameznikih. Seveda na račun naključnosti povečamo 'zasebnost' podatkov, a izgubimo pri natančnosti poizvedb, gre za ti. 'privacy-accuracy trade-off' (več o tem v nadaljevanju).

Naj bo $(\Omega, \mathcal{F}, \mathbb{P})$ verjetnostni prostor, $\mathbf{d} \in D^n$ opazovana podatkovna baza in $\mathcal{Q}(n)$ (n se nanaša na dimenzijo podatkovne baze) množica (možnih oz. dovoljenih) poizvedb. *Odzivni mehanizem* (za izbran nabor poizvedb $\mathcal{Q}(n)$) je potem definiran kot družina slučajnih spremenljivk

$$\{X_{Q,\mathbf{d}} : \Omega \to E_Q | Q \in \mathcal{Q}(n), \mathbf{d} \in D^n\}. \tag{1}$$

Pričakovana napaka takega mehanizma za dano poizvedbo Q in podatkovno bazo $\mathbf{d}$ je potem dana z $\mathbb{E}[\rho_Q(X_{Q,\mathbf{d}}, Q(\mathbf{d})]$. V nadaljevanju bo pogosto navedeno npr. $\mathbb{P}(X_{Q,\mathbf{d}} \in A)$, kar je seveda okrajšava za $\mathbb{P}(\{\omega \in \Omega : X_{Q,\mathbf{d}}(\omega) \in A\})$.

Ločimo dva glavna primera odzivnih mehanizmov:

- *Perturbacija podatkovne baze* (angl. sanitised response mechanism). Tu vnesemo naključnost v podatke, že preden podamo odgovor na poizvedbo. Za to potrebujemo družino merljivih preslikav (slučajnih vektorjev) $\{Y_{\mathbf{d}} : \Omega \to U^n | \mathbf{d} \in D^n\}$ Če taka družina obstaja, potem ima odzivni mehanizem obliko kompozituma.
$$X_{Q,\mathbf{d}} = Q \circ Y_{\mathbf{d}} \tag{2}$$
V praksi se ponavadi to izvede prek t. i. dodajanja šuma, torej $X_{\mathbf{d}} = \mathbf{d} + N$, kjer je N slučajni vektor v U^n. Za tak pristop (dodajanje šuma) je potrebno, da ima U^n primerno algebraično obliko (npr. vektorski prostor ali monoid, kar nam zagotovi zaprtost za seštevanje).

- *Perturbacija odgovorov na poizvedbo* (angl. output perturbation). Kot sklepamo že iz imena, tokrat podatke perturbiramo šele po poizvedbi. Recimo, da imamo podano poizvedbo $Q : U^n \to E_Q$. V primeru da obstaja družina merljivih preslikav $\{Z_q : \Omega \to E_Q | q \in E_Q\}$, je odzivni mehanizem definiran kot
$$X_{Q,\mathbf{d}} = Z_{Q(\mathbf{d})} \tag{3}$$

Opomba: Na tem mestu se nam naravno postavi vprašanje, kako postopamo pri konstrukciji odzivnega mehanizma v primeru funkcijskih podatkov. Dodajmo le, da bi tak opis presegal obseg tega diplomskega dela, saj je teorija zadaj, precej zahtevna in obsežna. Tu je poudarek bolj na opisu modela, ki omogoča enotno obravnavo različnih vrst podatkov, ne na sami konstrukciji odzivnih mehanizmov (čeprav bo v nadaljevanju podan primer na numeričnih podatkih).

3.3 Definicija diferencirane zasebnosti

Sedaj smo pripravili vse potrebno in lahko definiramo pojem diferencirane zasebnosti.

Definicija 1. Diferencirana zasebnost za posamezno poizvedbo Naj bo $\epsilon > 0$ in $0 \leq \delta \leq 1$. Odzivni mehanizem je (ϵ, δ)-diferencirano zaseben za poizvedbo Q, če za vse $\mathbf{d} \sim \mathbf{d'} \in D^n$ in za vse $A \in \mathcal{A}_Q$ velja

$$\mathbb{P}(X_{Q,\mathbf{d}} \in A) \leq e^{\epsilon}\mathbb{P}(X_{Q,\mathbf{d'}} \in A) + \delta \tag{4}$$

Opombe:

- Vidimo, da je diferencirana zasebnost koncept, ki se tiče odzivnega mehanizma (in ne npr. podatkov samih). Ideja zadaj je, da se odgovor na naš mehanizem ne sme preveč razlikovati za dve sosednji bazi (drugače povedano, če en posamezen vnos je ali ni v bazi, to ne bo preveč vplivalo na rezultat mehanizma). Na ta način diferencirano zasebni mehanizmi preprečujejo, da bi lahko prišli do sklepov o konkretnih posameznikih.

- Zasebnost opisujemo s parametroma ϵ in δ (na začetku se je uporabljal samo ϵ, t. i. stroga diferencirana zasebnost, vendar se je izkazalo, da je tak koncept pogosto preveč restriktiven za implementacijo v praksi; δ tako zajema verjetnost, da se osnovni mehanizem 'zlomi'). Nižja kot bosta oba parametra, večja bo zasebnost (a tudi manjša natančnost, bolj bodo odgovori na poizvedbe oddaljeni od resničnih). V praksi je tako izziv poiskati najmanjše možne vrednosti za ϵ in δ, pri kateri je natančnost še vedno dovolj visoka (gre za že omenjen 'trade-off between accuracy and privacy').

- Pomembna je tudi *simetričnost* definicije; neenakost mora veljati tudi če zamenjamo vlogi **d** in **d'**

Primer 5 Za boljše razumevanje ponazorimo definicijo še s primerom. Za dano poizvedbo Q izberimo konkretni sosednji bazi **d**, **d'** $\in D^n$ in $A \in \mathcal{A}_Q$. Označimo $a = \mathbb{P}(X_{Q,\mathbf{d}} \in A)$ in $b = \mathbb{P}(X_{Q,\mathbf{d'}} \in A)$. Zaradi simetričnosti morata v primeru (ϵ, δ)-diferencirane zasebnosti mehanizma, veljati obe neenakosti, torej $a \leq e^{\epsilon}b + \delta$ in $b \leq e^{\epsilon}a + \delta$. Obravnavajmo dva primera:

- a=b : Enakost obeh verjetnosti kaže na to, da se odgovor na poizvedbo ni spremenil z dodajanjem oz. odstranitvijo enega posameznika iz baze. Obe neenakosti tu sledita trivialno.

- a>b (brez škode za splošnost, zaradi simetričnosti nam ni treba obravnavati še primera a<b) : nenakost $b \leq e^{\epsilon}a + \delta$ tu prav tako sledi trivialno, druga neenakost pa nam podaja mejo, za koliko je lahko verjetnost b manjša od a, da bo obravnavani mehanizem še vedno diferencirano zaseben. V primeru $(\epsilon, \delta) = (0.05, 0.05)$,mehanizem, pri katerem bo za dan $A \in \mathcal{A}_Q$ $a = 0.9$ in $b = 0.8$, ne bo diferencirano zaseben, saj $0.9 \nleq e^{0.05}0.8 + 0.05 \doteq 0.89$ (ni izpolnjena druga neenakost).

Ta primer nakazuje tudi zahtevnost testiranja (ϵ, δ)-diferencirane zasebnosti mehanizma; zgornji postopek moramo namreč ponoviti za vse sosednje baze v D^n in za vse elemente A Borelove σ-algebra $\mathcal{A}_Q$ (teh je pogosto neštevno mnogo)! V praksi je to seveda v večini primerov neizvedljivo. V nadaljevanju bo podanih nekaj rezultatov, ki dane zahteve omilijo.

Definicija 2. Diferencirana zasebnost Odzivni mehanizem je (ϵ, δ)-diferencirano zaseben glede na $\mathcal{Q}(n)$ (množica poizvedb), če je (ϵ, δ)-diferencirano zaseben za vsako poizvedbo $Q \in \mathcal{Q}(n)$

4 Omilitev zahtev definicije diferencirane zasebnosti

4.1 Zadostne testne množice (angl. sufficient sets for differential privacy)

V prejšnjem poglavju smo definirali koncept diferencirane zasebnosti in izpostavili nekatere pomanjkljivosti. Ena izmed njih je bila, da je potrebno pogoj iz definicije (4) preveriti za vse elemente $\mathcal{A}_Q$ (σ-algebra množice možnih odgovorov na dano poizvedbo Q). Naslednji izrek nam pove, da je dovolj, da to preverimo le za vse elemente algebre S, ki $\mathcal{A}_Q$ generira.

Izrek 1 Naj bosta podana odzivni mehanizem (1) in poizvedba $(E_Q, \mathcal{A}_Q, Q)$. Naj bo $S \subset \mathcal{A}_Q$ algebra in naj velja $\sigma(\mathcal{S}) = \mathcal{A}_Q$. Če (4) velja za vse $A \in \mathcal{S}$, potem velja za vse $A \in \mathcal{A}_Q$.

Dokaz: Označimo z $\mathcal{B} \subset \mathcal{P}(E_Q)$ vse množice za katere je pogoj (4) izpolnjen. Po predpostavki iz izreka velja $\mathcal{S} \subseteq \mathcal{B}$. Naj bo $A_1 \subseteq A_2 \subseteq ..., A_i \in \mathcal{B}$ poljubno monotono naraščajoče zaporedje množic v $\mathcal{B}$. Označimo $\hat{A} = \bigcup_i A_i$. Naj bosta še **d**, **d'** $\in D^n$ poljubni sosednji podatkovni bazi. Potem velja

$$\begin{aligned}\mathbb{P}(X_{Q,\mathbf{d}} \in \hat{A}) = \lim_{i \to \infty} \mathbb{P}(X_{Q,\mathbf{d}} \in A_i) \leq \\ \leq e^{\epsilon} \lim_{i \to \infty} \mathbb{P}(X_{Q,\mathbf{d'}} \in A_i) + \delta = \\ = e^{\epsilon}\mathbb{P}(X_{Q,\mathbf{d}} \in \hat{A}) + \delta\end{aligned}$$

kjer smo pri obeh enakostih uporabili zveznost verjetnostne mere, pri neenakosti pa dejstvo, da za vsak i velja $A_i \in \mathcal{B}$. Identičen argument pokaže, da enako

velja za monotono padajoča zaporedja množic v $\mathcal{B}$. S tem smo pokazali, da je $\mathcal{B}$ monoton razred. Iz tega in iz dejstva, da je $\mathcal{S} \subseteq \mathcal{B}$ (po uporabi izreka o monotonih razredih), sledi, da je tudi $\sigma(\mathcal{S}) = \mathcal{A}_Q \subseteq \mathcal{B}$.

□

Primer 6 Uporabimo zgornji izrek na konkretnem zgledu. Vzemimo poizvedbo Q z $E_Q = C([0,1])$, torej zavzema vrednosti v prostoru zveznih funkcij. Kot normo vzemimo $||f||_\infty = sup\{|f(t)| : t \in [0,1]\}$ in pripadajočo σ-algebro $\mathcal{A}_Q$. Naj bo podan še odzivni mehanizem $X_{Q,\mathbf{d}}$. $X_{Q,\mathbf{d}}(\omega)$ torej leži v $C([0,1])$ za vsak $\omega \in \Omega$.
Izberimo še $k \in \mathbb{N}$ in k-terico realnih števil $0 \leq t_1 \ldots \leq t_k \leq 1$ in definirajmo preslikavo $\pi_{t_1,\ldots,t_k} : C([0,1]) \to \mathbb{R}^k$ kot

$$\pi_{t_1,\ldots,t_k}(f) = (f(t_1), \ldots, f(t_2)).$$

Te preslikave so merljive (natančneje $(\mathcal{A}_Q, \mathcal{B}(\mathbb{R}^k))$ merljive) in zato lahko definiramo $X_{Q,\mathbf{d}}^{t_1,\ldots,t_k} = \pi_{t_1,\ldots,t_k} \circ X_{Q,\mathbf{d}}$. Opazimo, da je sedaj $X_{Q,\mathbf{d}}^{t_1,\ldots,t_k}(\omega) \in \mathbb{R}^k$ za vsak $\omega \in \Omega$. Pokažimo, da če je končno-dimenzionalen mehanizem $X_{Q,\mathbf{d}}^{t_1,\ldots,t_k}$ diferencirano zaseben glede na definicijo (4), potem je tudi $X_{Q,\mathbf{d}}$ diferencirano zaseben. Najprej opazimo, da iz naše predpostavke sledi, da so opazovani mehanizmi zasebni za množice oblike $A = \pi^{-1}_{t_1,\ldots,t_k}(B)$, kjer je B Borelova množica v $\mathbb{R}^k$:

$$\begin{aligned}\mathbb{P}(X_{Q,\mathbf{d}} \in A) = \mathbb{P}(X_{Q,\mathbf{d}} \in \pi^{-1}_{t_1,\ldots,t_k}(B)) = \mathbb{P}(X_{Q,\mathbf{d}}^{t_1,\ldots,t_k} \in B) \leq \\ \leq e^\epsilon \mathbb{P}(X_{Q,\mathbf{d}}^{t_1,\ldots,t_k} \in B) + \delta = e^\epsilon \mathbb{P}(X_{Q,\mathbf{d}} \in A) + \delta\end{aligned}$$

Množice A tvorijo algebro na $C([0,1])$ (posledica lastnosti praslik) in izkaže se, da je σ-algebra, ki jo te množice generirajo, enaka Borelovi σ-algebri na $C([0,1])$, torej $\mathcal{A}_Q$ (dokaz tega na tem mestu izpustimo). Po uporabi zgornjega izreka sedaj sledi rezultat.

V tem zgledu smo problem preverjanja diferencirane zasebnosti za 'neskončno'-dimenzionalen mehanizem s pomočjo izreka prevedli na končno-dimenzionalen mehanizem. Rezultat je iz teoretičnega vidika zanimiv, nima pa pravega praktičnega pomena, saj je preverjanja pogoja na Borelovi σ-algebri $\mathbb{R}^k$ še vedno v praksi neizvedljiva naloga.

4.2 Identična poizvedba v primeru perturbacije podatkovne baze

Identična poizvedba je kot že ime pove, poizvedba, ki ne spremeni podatkovne baze. Odgovor na tako poizvedbo je torej celotna podatkovna baza, lahko bi rekli, da je identična poizvedba enaka javni objavi podatkov. Označimo jo z $(U^n, \mathcal{A}_{U_n}, I_n), I_n : D^n \to D^n$, $I_n(\mathbf{d})$=$\mathbf{d}$.

Ponavadi imamo opravka s sistemom, ki podpira več kot eno možno poizvedbo v podatkovno bazo. V tem primeru moramo pogoj diferencirane zasebnosti preveriti za vsako izmed razpoložljivih poizvedb posebej. Naslednji izrek pokaže, da je za mehanizme, ki perturbirajo podatkovno bazo (2), dovolj ta pogoj preveriti le za identično poizvedbo.

Izrek 2: Naj bo odzivni mehanizem s perturbacijo podatkovne baze (ϵ, δ)-diferencirano zaseben glede na identično poizvedbo $(U^n, \mathcal{A}_{U^n}, I_n)$. Potem sledi, da je tak mehanizem (ϵ, δ)-diferencirano zaseben glede na katerokoli poizvedbo $(E_Q, \mathcal{A}_Q, Q)$.

Opomba: Velja poudariti pomembnost tega, da v izreku ne postavimo nobenih omejitev na množico možnih odgovorov E_Q. Lahko bi bili naši podatki zelo enostavni, npr. naravna števila $D \in \mathbb{N}$, odgovori na poizvedbo pa bi bile funkcije ali zaporedja, tj. npr. $E_Q = C([0,1])$ ali $E_Q = l_\infty$. Zgornji izrek nam omogoči, da v tem primeru namesto da preverjamo pogoj (4) za vse elemente σ-algebre $C([0,1])$, moramo pogoj preveriti le za vse elemente σ-algebre $\mathbb{N}$ (saj je v primeru identične poizvedbe $E_Q = D$), kar je občutno lažje in lahko predstavlja razliko med v praksi izvedljivo in neizvedljivo nalogo!

Dokaz: Naj bosta $\mathbf{d}$, $\mathbf{d'} \in D^n$ poljubni sosednji podatkovni bazi. Po predpostavki velja

$$\mathbb{P}(Y_{\mathbf{d}} \in E) \leq e^{\epsilon}\mathbb{P}(Y_{\mathbf{d}'} \in E) + \delta \qquad (*)$$

za vsak $E \in \mathcal{A}_{U^n}$. Vzemimo sedaj poljubno poizvedbo $(E_Q, \mathcal{A}_Q, Q)$. Ker je $Q : U^n \to E_Q$ merljiva, velja $Q^{-1}(A) \in \mathcal{A}_{U^n}$ za vsak $A \in \mathcal{A}_Q$. Potem z uporabo (*) sledi

$$\begin{aligned}\mathbb{P}(X_{Q,\mathbf{d}} \in A) = \mathbb{P}(Q(Y_{\mathbf{d}}) \in A) &= \mathbb{P}(Y_{\mathbf{d}} \in Q^{-1}(A)) \\ &\leq e^{\epsilon}\mathbb{P}(Y_{\mathbf{d}'} \in Q^{-1}(A)) + \delta \\ &= \mathbb{P}(Q(Y_{\mathbf{d}'}) \in A) + \delta\end{aligned}$$

in vidimo, da je dan mehanizem diferenciano zaseben tudi za poljubno izbrano poizvedbo Q.

□

Sedaj lahko izpostavimo pomembno razliko med prej omenjenima vrstama odzivnih mehanizmov, perturbacija podatkovne baze (2) in perturbacija odgovorov na poizvedbe (3). Pri slednjem načinu gre ponavadi za to, da najprej izvedemo poizvedbo, šele na to pa prek dodajanja šuma zaščitimo podatke. V tem primeru obstaja možnost, da bi napadalec velikokrat izvedel identično poizvedbo in bi nato prek povprečja prišel do prave vrednosti iskanih podatkov. Posledično moramo omejiti število možnih poizvedb (oz. vrsto poizvedb, tj. prepovedati npr. identično poizvedbo). V primeru (2) pa tak napad ni mogoč, saj tu že preden podamo odgovor na poizvedbo zaščitimo podatke s primernim diferencirano zasebnim mehanizom. Odgovor na identično poizvedbo je tako vedno enak

in napadalec ne more priti do sklepov o pravi vrednosti podatkov v bazi.

Primer 7 Ponazorimo zgoraj povedano še na konkretnem primeru, ki je koristen tudi za boljše razumevanje formalizma opisanega modela ter razlike med posameznimi vrstami odzivnih mehanizmov. Recimo, da poizvedbo Q ponovimo k-krat ($k \geq 1$). To lahko modeliramo kot eno samo poizvedbo $Q^{(k)} : U^n \to E_Q^{(k)}$, kjer $E_Q^{(k)} = E_Q \times ... \times E_Q$. Torej je $Q^{(k)}(\mathbf{d}) = (Q(\mathbf{d}), ..., Q(\mathbf{d}))$. Iz zgornjega izreka vemo, da če je perturbacija podatkovne baze $Y_{\mathbf{d}}$ diferencirano zasebna, potem enako velja za poizvedbo $Q^{(k)} \circ Y_{\mathbf{d}}$ za poljuben k. Res vidimo, da možnost večkratnih poizvedb mehanizme te vrste ne ogroža.

Enako pa ne velja za mehanizme, ki perturbirajo odgovore. Večkratne poizvedbe tu namreč lahko vodijo do zloma diferencirano zasebnih sistemov. Vzemimo preprosto poizvedbo $Q : U^n \to \{0, 1\}$, torej $E_Q = \{0, 1\}$. Da definiramo odzivni mehanizem odgovorov na to poizvedbo, navedimo porazdelitvi Z_0 in Z_1. Če nastavimo $\mathbb{P}(Z_i = i) = 1 - p$ in $\mathbb{P}(Z_i \neq i) = p$ za $i = 0, 1$, potem je mehanizem $X_{Q,\mathbf{d}} = Z_{Q(\mathbf{d})}$ (ϵ, δ) diferencirano zaseben natanko tedaj, ko velja

$$p \geq \frac{1 - \delta}{1 + e^{\epsilon}}$$

(brez izgube za splošnost dodatno privzamemo, da je $p < \frac{1}{2}$). Naravno lahko predpostavimo, da obstajata sosednji podatkovni bazi **d**, **d'** v D^n za katere bo odgovor na poizvedbo Q različen, npr. $Q(\mathbf{d}) = 0, Q(\mathbf{d'}) = 1$. Potem za množico $A = \{0\}$ velja $\mathbb{P}(Z_{Q(\mathbf{d})} \in A) = \mathbb{P}(Z_0 = 0) = 1 - p$ in podobno $\mathbb{P}(Z_{Q(\mathbf{d'})} \in A) = p$. Če torej vzamemo $\epsilon = 0.1$ in $\delta = 0.4$ ter nastavimo $p = 0.286$, bo opisan mehanizem diferencirano zaseben za poizvedbo Q.

Recimo, da sedaj poizvedbo Q uporabimo dvakrat. Množica možnih odgovorov je torej $E_Q \times E_Q$, oz. konkretno $\{0, 1\} \times \{0, 1\}$. Slučajne spremenljivke $Z_{(q_1,q_2)}$ za $q_1, q_2 \in E_Q$ najlažje definiramo kot $Z_{(q_1,q_2)} = (Z_1, Z_2)$, kjer sta Z_i neodvisni in enako kot Z_{q_i} porazdeljeni slučajni spremenljivki za $i = 1, 2$. Ob isti izbiri p dobimo

$$\mathbb{P}(Z^{(2)}_{Q(\mathbf{d})} \in A \times A) = (1 - p)^2 = 0.5098$$

ter

$$\mathbb{P}(Z^{(2)}_{Q(\mathbf{d'})} \in A \times A) = p^2 = 0.0817.$$

Očitno velja

$$\mathbb{P}(Z^{(2)}_{Q(\mathbf{d})} \in A \times A) > e^{\epsilon}\mathbb{P}(Z^{(2)}_{Q(\mathbf{d'})} \in A \times A) + \delta.$$

Tako vidimo, da je dvakratna uporaba poizvedbe Q 'zlomila' mehanizem $Z_{Q(\mathbf{d})}$.

4.3 Poenostavitev na 1D baze (brez dokaza)

V tem poglavju bomo vpeljali posebno obliko odzivnih mehanizmov, ki nam omogoča, da pogoj diferencirane zasebnosti preverjamo le za 1-dimenzionalne

baze (tj. imamo le en vnos oz. enega posameznika). Očitno je, da ta lastnost precej olajša testiranje. V splošnem moramo namreč preveriti pogoj za vse n-dimenzionalne sosednje baze.

Predpostavimo, da obstaja družina slučajnih spremenljivk (merljivih preslikav) oblike $\{Y_d : \Omega \to U | d \in D\}$. Potem za $\mathbf{d} = (d_1, ..., d_n)$ definiramo odzivni mehanizem $Y_{\mathbf{d}}$ kot

$$Y_{\mathbf{d}}(\omega) = (Y_{d_1}(\omega), ..., Y_{d_n}(\omega)), \tag{5}$$

kjer so Y_{d_i} med sabo neodvisne. To nam zagotavlja obstoj marginalnih porazdelitev. Lahko so Y_{d_i} tudi enako porazdeljene slučajne spremenljivke, ni pa to nujno. Torej je $Y_{\mathbf{d}}$ n-dimenzionalen mehanizem, sestavljen po komponentah iz n 1-dimenzionalnih mehanizmov, ki so med seboj neodvisni.

Opomba: Dodajmo še, da je zgoraj definiran mehanizem, le posebna oblika perturbacije podatkovne baze (2) (tu dodatno zahtevamo obstoj marginalnih porazdelitev). To nam omogoča, da v spodnjem izreku delamo le z identično poizvedbo, torej $Y_{Q,\mathbf{d}} = Y_{\mathbf{d}}$. Za vse ostale poizvedbe potem rezultat sledi kot posledica izreka v prejšnjem poglavju.

Izrek 3 Naj bo podana družina 1-dimenzionalnih diferencirano zasebnih mehanizmov $\{Y_d : \Omega \to U | d \in D\}$. Velja torej

$$\mathbb{P}(Y_d \in A) \leq e^{\epsilon}\mathbb{P}(Y_{d'} \in A) + \delta$$

za vse $d, d' \in D, A \in \mathcal{A}_D$. Če definiramo n-dimenzionalni odzivni mehanizem kot (5), potem sledi, da je tudi ta diferencirano zaseben:

$$\mathbb{P}(Y_{\mathbf{d}} \in A) \leq e^{\epsilon}\mathbb{P}(Y_{\mathbf{d'}} \in A) + \delta$$

za vse $\mathbf{d} \sim \mathbf{d'} \in D^n, A \in D^n$.

Dokaz tu izpustimo (vsebuje veliko tehničnih detajlov). Bralec si ga lahko ogleda v izhodiščnem članku.

Primer 8: Zgornji izrek je še posebej uporaben v primeru, ko je D diskreten metrični prostor. Ponazorimo to s primerom. Naj bo D končen prostor z |D| elementi. Za mehanizme oblike (5) moramo tako preveriti pogoj diferencirane zasebnosti le za $\binom{|D|}{2}$ elementov D (število sosednjih podatkovnih baz v 1-dimenzionalnem primeru je kar enako številu možnih parov) in za $2^{|D|}$ podmnožic D (število elementov v $\mathcal{A}_D$). To nam že zagotavlja diferencirano zasebnost tudi v primeru D^n. Brez izreka bi morali pogoj testirati za $n\binom{|D|}{2}|D|^{n-1}$ elementov D^n (število sosednjih baz v n-dimenzionalnem primeru) in za $2^{|D|^n}$ podmožic (število elementov v $\mathcal{A}_{D^n}$).

Primer 9 Oglejmo si enostaven primer diferencirano zasebnega odzivnega mehanizma: *Laplaceov odzivni mehanizem za numerične podatke.* Naj bodo naši

podatki elementi $D \subset \mathbb{R}$. Predpostavimo tudi, da je D omejen, kar v nadaljevanju potrebujemo za obstoj diam(D). $L : \Omega \to \mathbb{R}$ naj bo Laplacovo porazdeljena slučajna spremenljivka s parametroma $(0, b), b > 0$. Verjetnostna gostota ima potem obliko $f(x) = \frac{1}{2b} e^{-\frac{|x|}{b}}$. Za vsak $d \in D$ potem definirajmo 1-dimenzionalni mehanizem kot $Y_d(\omega) = d + L(\omega)$. Parameter b izberimo tako, da

$$b \geq \frac{diam(D)}{\epsilon - \log(1 - \delta)}.$$

Potem sledi, da je vsak n-dimenzionalen mehanizem oblike (5) (ϵ, δ) diferencirano zaseben za vsako n-dimenzionalno podatkovno bazo D^n in vsako poizvedbo. To lahko enostavno pokažemo.

Najprej vemo, da $\mathbb{P}(Y_d(\omega) \in A) = \mathbb{P}(d + L(\omega) \in A) = \int_A \frac{1}{2b} e^{\frac{-|x-d|}{b}} dx$. Kot posledica izrekov 2 in 3 potem vemo, da bo zgornja trditev veljala natanko tedaj, ko bo veljajo

$$\int_A \frac{e^{\frac{-|x-d|}{b}}}{2b} dx \leq e^{\epsilon} \int_A \frac{e^{\frac{-|x-d'|}{b}}}{2b} dx + \delta$$

za vse $d, d' \in D, A \in \mathcal{B}(\mathbb{R})$. Zgornja neenakost pa bo veljala natanko tedaj, ko velja $1 \leq e^{\epsilon - \frac{|d-d'|}{b}} + \delta$ (uporabimo trikotniško neenakost $|x-d'| \leq |x-d| + |d-d'|$ in dejstvo, da nam neenakost porodi najstrožji pogoj v primeru $A = \mathbb{R}$). Po preureditvi sledi rezultat.

5 Natančnost diferencirano zasebnih mehanizmov

Zaenkrat smo se posvetili vprašanju zasebnosti odzivnih mehanizmov, v tem poglavju pa se bomo vprašali še o njihovi natančnosti. Delali bomo z mehanizmi oblike (5). Ker so slednji zgrajeni iz 1-dimenzionalnih mehanizmov, se bomo tu osredotočili na njihovo natančnost. Te rezulate lahko potem uporabimo za izpeljavo napake v n-dimenzionalnem primeru, torej na D^n, točna oblika pa bo odvisna od metrike ρ_n.

Za izbrani $d \in D$ vemo, da je metrika $\rho(\cdot, d)$ zvezna funkcija (sledi iz trikotniške neenakosti). Torej je taka funkcija tudi Borelovo merljiva (natančneje Borel - Borelovo merljiva) na D. Sledi, da je $\rho(Y_d, d)$ nenegativna slučajna spremenljivka (kompozitum merljivih funkcij je merljiva funkcija). Sedaj lahko definiramo maksimalno pričakovano napako γ danega mehanizma Y_d:

$$\gamma := max_{d \in D} \mathbb{E}[\rho(Y_d, d)].$$

Za dani $r > 0$ in $x \in D$ označimo z $B_r(x)$ odprto kroglo.

Lema 1: Naj bo podana družina 1-dimenzionalnih diferencirano zasebnih mehanizmov $\{Y_d : \Omega \to U | d \in D\}$ (glej definicijo 1) in naj velja $0 \leq \delta < 1$. Potem sledi $\gamma > 0$.

Opomba: δ je tu strogo manjša od 1. V primeru da je $\delta = 1$, je namreč vsak mehanizem diferencirano zaseben.

Dokaz: Ker je D kompakten, lahko izberemo $u, v \in D$ z $\rho(u, v) = diam(D)$. Definirajmo $r = \frac{diam(D)}{2}$. Če obrnemo pogoj diferencirane zasebnosti dobimo

$$\mathbb{P}(Y_u \in B_r(v)) \geq e^{-\epsilon}(\mathbb{P}(Y_v \in B_r(v)) - \delta).$$

Ker je $\rho(x, u) \geq r > 0$ za vsak $x \in B_r(v)$, sledi

$$\begin{aligned}\mathbb{E}[\rho(Y_u, u)] &\geq \mathbb{E}[\rho(Y_u, u) | Y_u \in B_r(v)]\mathbb{P}(Y_u \in B_r(v)) \\ &\geq re^{-\epsilon}(\mathbb{P}(Y_v \in B_r(v)) - \delta) > 0.\end{aligned}$$

Posebej moramo obravanati le še primer, ko je $\delta = \mathbb{P}(Y_v \in B_r(v))$ (brez škode za splošnost namreč lahko privzamemo $\delta \leq \mathbb{P}(Y_v \in B_r(v))$, saj drugače pogoj diferencirane zasebnosti ne bi imel smisla). Rezultat v tem primeru sledi po podobnem premisleku kot zgoraj.

□

Spodnja izreka podata spodnjo mejo napake pri (ϵ, δ)-zasebnih mehanizmih. Pri obeh opazimo, da večja kot je zasebnost (tj. manjša kot sta ϵ in δ), višja je spodnja meja napake (tj. naš mehanizem je manj natančen). To se sklada z našo definicijo, pri kateri smo omenil t. i. 'privacy-accuracy trade-off'.

Izrek 4 Naj bo podana družina 1-dimenzionalnih diferencirano zasebnih mehanizmov $\{Y_d : \Omega \to U | d \in D\}$ (glej definicijo 1). Potem velja

$$\gamma \geq (1 - \delta)(\frac{diam(D)}{2(1 + e^{\epsilon})})$$

Dokaz: D je tu ponovno kompakten metričen prostor iz česar sledi, da obstajata točki $u, v \in D$ z $\rho(u, v) = diam(D)$. Iz zgornje leme vemo, da je $\gamma > 0$. Tako lahko definiramo $t := \frac{diam(D)}{2\gamma}$. Sedaj opazimo, da sta odprti krogli $B_{t\gamma}(u), B_{t\gamma}(v)$ disjunktni. Z uporabo neenakosti Markova na pozitivni slučajni spremenljivki $\rho(Y_u, u)$ dobimo:

$$\mathbb{P}(Y_u \in B_{t\gamma}(u)) = \mathbb{P}(\rho(Y_u, u) < t\gamma) \geq 1 - \frac{\mathbb{E}(\rho(Y_u, u))}{t\gamma} = 1 - \frac{2\gamma}{diam(D)}$$

Iz tega nato sledi:

$$\mathbb{P}(Y_u \in B_{t\gamma}(v)) \leq \mathbb{P}(\{Y_u \in B_{t\gamma}(u)\}^C) \leq \frac{2\gamma}{diam(D)}$$

Hkrati vemo (pogoj diferencirane zasebnosti)

$$\mathbb{P}(Y_u \in B_{t\gamma}(v)) \geq e^{-\epsilon}(\mathbb{P}(Y_v \in B_{t\gamma}(v)) - \delta),$$

kar nam po uporabi prejšnjih dveh neenakosti, da sledeče:

$$\frac{2\gamma}{diam(D)} \geq e^{-\epsilon}(1 - \frac{2\gamma}{diam(D)} - \delta)$$

Sedaj izrazimo γ in dobimo željeno spodnjo mejo.

□

V zgornjem izreku je bil D poljuben kompakten metrični prostor. Sedaj predpostavimo dodatno še, da je D diskreten metrični prostor, kar pomeni, da obstaja $\kappa > 0$, da velja

$$\rho(u, v) \geq \kappa \quad \forall u, v \in D.$$

Navedimo še lemo, ki jo potrebujemo v spodnjem izreku.

Lema: Naj bo (D, ρ) diskreten metričen prostor. Če je D kompakten, potem je D končen (angl. finite).

Opomba: V zgornji lemi smo navedli implikacijo, čeprav v resnici velja ekvivalenca med kompaktnostjo in končnostjo. Za potrebo tega diplomskega dela obratne smeri ne potrebujemo.

Dokaz: Predpostavimo, da je D neskončen (angl. infinite). Potem obstaja neskončno zaporedje različnih elementov iz D. Iz kompaktnosti sledi, da ima tako zaporedje konvergentno podzaporedje z limito v D. Vemo, da so v diskretnem prostoru edina konvergentna zaporedja konstantna (okoli vsake točke v takem prostoru lahko najdemo kroglo, v kateri ne leži nobena druga točka). Nobeno podzaporedje zaporedje različnih elementov ni konstantno, kar bi pomenilo, da D ni kompakten in na ta način smo prišli do protislovja. Sledi, da je D končnen metrični prostor.

Izrek 5 Naj bo D diskreten metričen prostor z $|D| = m+1$ in $\kappa = min_{d,d' \in D}\rho(d, d')$. Naj bo podana družina 1-dimenzionalnih diferencirano zasebnih mehanizmov $\{X_d : \Omega \to U | d \in D\}$ (glej definicijo 1). Potem velja

$$\gamma \geq (1 - \delta)(\frac{\kappa m}{m + e^{\epsilon}})$$

Dokaz: Najprej definiramo $t := \frac{\kappa}{\gamma}$. Opazimo, da je $m + 1$ odprtih krogel $B_{t\gamma}(u), u \in D$ disjunktnih. Sedaj izberemo točno določen $u \in D$ in po enakem razmisleku kot v zgornjem izreku po uporabi neenakosti Markova sledi

$$\mathbb{P}(X_u \in B_{t\gamma}(u)) \geq 1 - \frac{\gamma}{\kappa}.$$

Dodatno mora obstajati tak $v \neq u$, da velja

$$\mathbb{P}(X_u \in B_{t\gamma}(v)) \leq \frac{\gamma}{\kappa m}.$$

Izberemo tak v in uporabimo pogoj diferencirane zasebnosti (enako kot v zgornjem dokazu), kar nam po uporabi obeh zgornjih neenakosti, da

$$\frac{\gamma}{\kappa m} \geq e^{-\epsilon}(1 - \frac{\gamma}{\kappa} - \delta).$$

Sedaj izrazimo γ in dobimo željeno spodnjo mejo.

□

Opomba: Vidimo, da po tem izreku za velike m-je dobimo oceno spodnje meje približno $(1 - \delta)\kappa$. Spodnja meja po tem izreku je tako boljša (torej nižja), v primeru ko velja $min_{d,d' \in D}\rho(d, d') = \kappa < \frac{diam(D)}{2(1+e^{\epsilon})} = \frac{max_{d,d' \in D}\rho(d,d')}{2(1+e^{\epsilon})}$. Primer takega prostora je npr. prostor prvih 100 naravnih števil opremljen z evklidsko metriko.