UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 1. stopnja

Metod Jazbec

Splošna definicija diferencirane zasebnosti

Delo diplomskega seminarja

Mentor: doc. dr. Aljoša Peperko

Ljubljana, 2018

Kazalo

Splošna definicija diferencirane zasebnosti

Povzetek

V delu predstavimo diferencirano zasebnost. Gre za matematično definicijo zasebnosti pri javni objavi ter rudarjenju podatkov. Predstavljena je splošna definicija v kontekstu metričnih prostorov in verjetnostne mere, ki omogoča enotno obravnavo različnih vrst podatkov. Pokažemo nekaj osnovnih izrekov, ki omilijo zahteve definicije. Obravnavan je Laplacov mehanizem za numerične podatke. Podana je izpeljava spodnjih mej za največjo napako zasebnih odzivnih mehanizmov. V nadaljevanju se osredotočimo na funkcijske podatke. S pomočjo teorije Gaussovih procesov in Hilbertovih prostorov z reprodukcijskim jedrom pokažemo uporabo diferencirane zasebnosti na primeru jedrne cenilke gostote. Osnovne mehanizme implementiramo in predstavimo rezultate.

General definition of differential privacy

Abstract

We introduce the concept of differential privacy, mathematical definition for privacy preserving data publishing and data mining. General definition in context of metric spaces and probability measure is given. Further, we present some theorems which help to alleviate the requirements of described definition. Laplace mechanism for numerical data and lower bounds on errors of response mechanisms are presented. We later turn focus to functional data. Using Gaussian process and Reproducing Kernel Hilbert Spaces we present how differential privacy is used for privatization of density kernel estimator. Most of the described mechanism are also implemented and results are presented at the end.

Math. Subj. Class. (2010): navedite vsaj eno klasifikacijsko oznako – dostopne so na `www.ams.org/mathscinet/msc/msc2010.html`

Ključne besede: diferencirana zasebnost, odzivni mehanizem, metrični prostor, funkcijski podatki, verjetnostna mera

Keywords: differential privacy, response mechanism, metric space, functional data, probability measure

1. Uvod

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 zasebnosti? Obstaja veliko različnih pristopov (najbolj pogoste so t. i. 'metode anonimizacije'), vendar se je izkazalo, da je veliko od teh neprimernih in so izpostavljeni raznim 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 probabilistični. 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 [3,4]. 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 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 (angl. data publishing) in pridobivanja podatkov (angl. 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 tudi izkaže, da lahko v primeru determinističnih 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

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

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

in iz razlike obeh sklepal, da Edi ima diabetes. Odgovor na to problematiko ponuja diferencirana zasebnost, pri kateri poizvedbe postanejo 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 [3], ki nekatere podatke še preden jih iz naprave (npr. iz iphona) pošlje na centralni server, 'zamaskira' s pomočjo algoritmov, 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

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

Definicija 2.1. 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 številne 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 številne preseke).

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

Izrek 2.2. *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 $B \in \mathcal{S}$ skupaj implicirata $A \cup B \in m(\mathcal{S})$. Povedano drugače, $B \in S$ implicira $B \in \mathcal{H}_2$. Torej je $\mathcal{S} \subseteq \mathcal{H}_2$ 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 o 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 3.1.
- 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 2.3. 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 2.4. Zgoraj podana definicija ne opisuje najbolj splošnega pojma kompaktnosti, ampak gre za t. i. kompaktnost glede na zaporedja (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.

Lema 2.5. *Naj bo (D, ρ) končen metričen prostor. Potem je D kompakten.*

Dokaz. 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.

Definicija 2.6. Naj bo D poljubna množica in $a, b \in D^n$. Potem je *Hammingova razdalja* $\rho_H(a, b)$ enaka številu mest, na katerih se vektorja a in b razlikujeta.

Opazimo, da je v 1-dimenzionalnem primeru 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. Hammingova 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.

V zadnjem poglavju bomo za izpeljavo nekaterih rezultatov potrebovali še osnovno teorijo jedrnih funkcij v statistični analizi, Gaussovih procesov ter Hilbertovih prostorov z reprodukcijskim jedrom. Z namenom večje preglednosti navedemo še izrek Radon-Nikodym.

Definicija 2.7. Funkcija K je jedrna funkcija (angl. kernel function), če je nenegativna, integrabilna ter slika v prostor realnih števil.

Opomba 2.8. Zgoraj podana definicija je najbolj splošna, ponavadi pa zahtevamo še dva dodatna pogoja in sicer simetričnost, torej $K(x) = K(-x)$, ter normalizacijo

$$\int_{-\infty}^{\infty} K(u)du = 1.$$

Enostaven primer je eno-dimenzionalno Gaussovo jedro, ki ima obliko $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ (vidimo, da gre kar za gostoto standarno normalne porazdelitve).

Definicija 2.9. Gaussov proces, parametriziran z indeksno množico T, je slučajni proces $\{X_t : t \in T\}$, za katerega velja, da je za vsak končni nabor točk $t_1, ..., t_n \in T$ slučajni vektor

$$(X_{t_1}, ..., X_{t_n})$$

porazdeljen multivariatno normalno.

Opomba 2.10. Gaussov proces je kot slučajni proces definiran na nekem verjetnostnem prostoru $(\Omega, \mathcal{F}, \mathbb{P})$. Pri fiksnem $\omega \in \Omega$ je preslikava $T \to \mathbb{R}$ (oz. $t \to X_t(\omega)$) realizacija ali trajektorija Gaussovega procesa X_t. Gaussov proces je popolnoma določen s funkcijama povprečja in kovariance:

$$m(t) = \mathbb{E}X_t, \quad K(s, t) = Cov(X_s, X_t).$$

Spomnimo se, da je Hilbertov prostor $\mathcal{H}$ poln metrični prostor (vsako Cauchyjevo zaporedje ima limito vsebovano v tem prostoru), kjer je metrika podana prek skalarnega produkta, torej za $f, g \in \mathcal{H}$ velja $d(f, g) = \sqrt{< f - g, f - g >}$.

Definicija 2.11. Naj bo H Hilbertov prostor, katerega elementi so funkcije oblike $f : X \to \mathbb{R}$ (pri tem je X poljubna množica). Označimo z L_x linearen funkcional, ki vsako funkcijo $f \in \mathcal{H}$ izvrednoti v x, torej

$$L_x : f \to f(x).$$

Potem je $\mathcal{H}$ Hilbertov prostor z reprodukcijskim jedrom natanko tedaj, ko je L_x zvezen operator nad $\mathcal{H}$ za vsak $x \in X$.

Opomba 2.12. Opišimo podano definicijo ter reprodukcijsko lastnost na še bolj intiutiven način. Z uporabo Rieszovega reprezentacijskega izreka o funkcionalih opazimo, da za vsak $x \in X$ obstaja natanko en $K_x \in H$ z lastnostjo

$$f(x) = L_x(f) = \langle f, K_x \rangle_{\mathcal{H}}, \quad \forall f \in H.$$

Če sedaj namesto f vzamemo $K_x \in H$, sledi, da za vsak $y \in X$ obstaja $K_y \in H$, da velja

$$K_x(y) = L_y(K_x) = \langle K_y, K_x \rangle_{\mathcal{H}}.$$

To nam omogoča, da definiramo reprodukcijsko jedro Hilbertovega prostora $\mathcal{H}$ kot funckijo $K : X \times X \to \mathbb{R}$, kjer je $K(x, y) = \langle K_y, K_x \rangle_{\mathcal{H}}$.

Alternativno si lahko Hilbertov prostor $\mathcal{H}$ predstavljamo kot zaprtje linearne ogrinjače funckij oblike $K_x, x \in X$. Za dve funkciji oblike $f = \sum_{i=1}^{n} \theta_i K_{x_i}$ in $g = \sum_{j=1}^{m} \xi_j K_{y_j}$ $(\xi_i, \theta_i \in \mathbb{R} \forall i, x_1, ..., x_n, y_1, ..., y_m \in X)$ ima tako njun skalarani produkt obliko

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{m} \theta_i \xi_j K(x_i, y_j).$$

Definicija 2.13. Naj bosta μ in ν meri definirani na istem merljivem prostoru (X, Σ). Rečemo, da je μ absolutno zvezna glede ne ν (tudi da ν dominira μ), če velja $\mu(A) = 0$ za vse merljive množice A, za katere velja $\nu(A) = 0$. Označimo $\mu \ll \nu$.

Izrek 2.14 (Radon-Nikodym)**.** *Naj bosta μ in ν meri definirani na istem merljivem prostoru (X, Σ). Dodatno naj velja, da sta μ in ν σ-končni meri. Če velja $\mu \ll \nu$, potem obstaja merljiva funkcija $f : X \to [0, \infty)$, da za vsako merljivo množico $A \subseteq X$ velja*

$$\mu(A) = \int_A f d\nu = \int_A \frac{d\mu}{d\nu} d\nu$$

Funkcija f se imenuje Radon-Nikodymjev odvod.

Opomba 2.15. V verjetnosti je gostota slučajne spremenljivke X s porazdelitvenim zakonom F_X ravno Radon-Nikodymjev odvod inducirane mere dF_X glede na neko osnovno mero (v primeru zveznih slučajnih spremenljivk ponavadi vzamemo Lebesgueovo mero λ, torej dobimo $f = \frac{dF_X}{d\lambda}$. Včasih rečemo, da je slučajna spremenljivka X absolutno zvezna, če ima gostoto (v Radon-Nikodymjevem smislu).

3. Splošni podatkovni model in definicija diferencirane zasebnosti

3.1. **Predstavitev podatkovne baze.** 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 (gre za 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 najmanjša σ-algebra, ki vsebuje $\{A_1 \times ... \times A_n | A_i \in \mathcal{A}_U\}$ 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 3.1. 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 enaka $2^{2^{\mathcal{H}}}$. 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 3.2. 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 produktno σ-algebro, torej $\mathcal{A}_D = \sigma(\{A_1 \times A_2 \times A_3 | A_1 \in \mathcal{B}(\mathbb{R}^n), A_2 \in \mathcal{B}(\mathbb{R}^m), A_3 \in \mathcal{B}(\mathbb{R}^3)\})$. Ta koncept lahko razširimo na 3-D barvne slike in tudi na video posnetke. ♢

Primer 3.3. 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

$$D = \{1, 2, ..., st.pacientov\} \times \{1, ..., 120\} \times \{M, F\} \times \\ \{Ljubljana, ..., SpodnjiDuplek\} \times \{0, 1\}^n.$$

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

Primer 3.4. Navedimo še primer, ko imamo opravka s t. i. funkcijskimi podatki. Ti 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 prostori), kar pomeni, da imajo naravno podano normo. Le-ta nam inducira metriko ρ, prav tako pa lahko prek norme pridemo do 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, je $\mathbf{d} \sim \mathbf{d}'$ natanko tedaj, ko je $\rho_H(\mathbf{d},\mathbf{d}')= 1$.

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 maksimuma v tej definiciji je posledica dejstva, da zvezna funkcija ρ doseže svoj maksimum na kompaktni množici D.

3.2. **Modeliranje poizvedb.** Poizvedba (angl. 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 $A \in \mathcal{A}_Q$.

Primer 3.5. 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 ($E_Q = 2^{\mathcal{H}}$). $\diamondsuit$

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 t. i. 'privacy-accuracy trade-off' (več o tem v nadaljevanju).

Definicija 3.6. 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 **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, še 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 $Y_{\mathbf{d}} = \mathbf{d} + N$, 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

$$X_{Q,\mathbf{d}} = Z_{Q(\mathbf{d})}. \tag{3}$$

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

Definicija 3.7. 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

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

Opomba 3.8.

- 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 $\mathbf{d}$ in $\mathbf{d'}$

Primer 3.9. Za boljše razumevanje ponazorimo definicijo še s primerom. Za dano poizvedbo Q izberimo konkretni sosednji bazi $\mathbf{d}$, $\mathbf{d'} \in D^n$ 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 $(\epsilon, \delta) = (0.05, 0.05)$, (ϵ, δ)-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 3.10. 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.** 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 pogoj (4) preverimo le za vse elemente algebre $\mathcal{S}$, ki $\mathcal{A}_Q$ generira.

Izrek 4.1. *Naj bosta podana odzivni mehanizem* (1) *in poizvedba* $(E_Q, \mathcal{A}_Q, Q)$. *Naj bo* $\mathcal{S} \subset \mathcal{A}_Q$ *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 $A_1 \subseteq A_2 \subseteq ...$, kjer je $A_i \in \mathcal{B}$ za vsak $i \in \mathbb{N}$, poljubno monotono naraščajoče zaporedje množic v $\mathcal{B}$. Naj bosta še **d**, **d'** $\in D^n$ poljubni sosednji podatkovni bazi. Potem velja

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

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 4.2. 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 $0 \leq t_1 \leq \ldots \leq t_k \leq 1$ in definirajmo preslikavo $\pi_{t_1,...,t_k} : C([0, 1]) \to \mathbb{R}^k$ kot

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

Te preslikave so merljive (natančneje $(\mathcal{A}_Q, \mathcal{B}(\mathbb{R}^k))$ merljive) in zato lahko definiramo $X^{t_1,\ldots,t_k}_{Q,\mathbf{d}} = \pi_{t_1,\ldots,t_k} \circ X_{Q,\mathbf{d}}$. Opazimo, da je sedaj $X^{t_1,\ldots,t_k}_{Q,\mathbf{d}}(\omega) \in \mathbb{R}^k$ za vsak $\omega \in \Omega$. Pokažimo, da če je končno-dimenzionalen mehanizem $X^{t_1,\ldots,t_k}_{Q,\mathbf{d}}$ diferencirano zaseben glede na definicijo (4) za vse k-terice $0 \leq t_1 \leq \ldots \leq t_k \leq 1$, 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 $\mathbb{R}^k$:

$$\mathbb{P}(X_{Q,\mathbf{d}} \in A) = \mathbb{P}(X_{Q,\mathbf{d}} \in \pi^{-1}_{t_1,\ldots,t_k}(B)) = \mathbb{P}(X^{t_1,\ldots,t_k}_{Q,\mathbf{d}} \in B) \leq$$
$$\leq e^{\epsilon}\mathbb{P}(X^{t_1,\ldots,t_k}_{Q,\mathbf{d'}} \in B) + \delta = e^{\epsilon}\mathbb{P}(X_{Q,\mathbf{d'}} \in A) + \delta, \forall \mathbf{d} \sim \mathbf{d'} \in D^n$$

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 $\mathcal{A}_Q$ (glej izrek 7.2.1 v [9]). 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 4.3. *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 4.4. 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 $2^{\mathbb{N}}$ (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

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

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

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

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 če 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 4.5. 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 $\mathbf{d}$, $\mathbf{d'}$ 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 $Z_{Q(\mathbf{d})}$. $\diamondsuit$

4.3. **Poenostavitev na eno-dimenzionalne baze.** 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 iz n eno-dimenzionalnih mehanizmov, ki so med seboj neodvisni.

Opomba 4.6. 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 4.7. *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 v (5)*, potem sledi, da je tudi ta diferencirano zaseben:*

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

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

Dokaz tu izpustimo (vsebuje veliko tehničnih detajlov). Bralec si ga lahko ogleda v izhodiščnem članku [1].

Primer 4.8. 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 4.9. 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 eno-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. ♢

Primer 4.10. Navedimo še primer mehanizma za diskretne podatke. Navežimo se na primer 3.1 in recimo, da $D = 2^{\mathcal{H}}$ predstavlja množico vseh možnih hobijev. Kot prej, predpostavimo končnost D, torej $|D| = m + 1$ za nek $m \in \mathbb{N}$. Po vzoru izreka 4.7 bomo skonstruirali mehanizem za eno-dimenzionalne baze, ki ga lahko nato brez težav prenesemo v n-dimenzionalen primer z uporabo (5).

Za $d \in D$ definirajmo diskretno slučajno spremenljivko Y_d (zavzame $|D|$ možnih vrednosti) prek naslednje verjetnostne funkcije:

$$\mathbb{P}(Y_d = d) = 1 - pm, \mathbb{P}(Y_d = d') = p.$$

Tu je $d \neq d' \in D$. Naravno predpostavimo $1 - pm > p$, kar pomeni, da bomo z večjo verjetnostjo podali pravilen odgovor na poizvedbo (v nasprotnem primeru nam tudi še tako močna zasebnost ne koristi, saj izgubimo preveč natančnosti).

Da bo podan mehanizem (ϵ, δ)-diferencirano zaseben mora veljati:

$$\mathbb{P}(Y_d \in A) \leq \mathbb{P}(Y_{d'} \in A)e^{\epsilon} + \delta \tag{1}$$

za vsak $A \subset D$ in $d, d' \in D$.

Pokažimo, da bo dan mehanizem zadoščal pogojem diferencirane zasebnosti natanko tedaj, ko bo

$$1 - pm \leq pe^{\epsilon} + \delta.$$

Ta pogoj je zagotov potreben, kar vidimo, če za A vzamemo enostavno množico $\{d\}$. Da pokažemo, da je ta pogoj tudi zadosten, moramo obravnavati 4 različne primere.

(1) $d, d' \notin A$: Velja $\mathbb{P}(Y_d \in A) = \mathbb{P}(Y_{d'} \in A) = p|A|$ in pogoj diferencirane zasebnosti sledi trivialno.
(2) $d, d' \in A$: Velja $\mathbb{P}(Y_d \in A) = \mathbb{P}(Y_{d'} \in A) = p(|A| - 1) + 1 - pm = p(|A| - m - 1) + 1$. Tudi tu pogoj diferencirane zasebnosti sledi trivialno.
(3) $d \notin A, d' \in A$: Velja $\mathbb{P}(Y_d \in A) \leq \mathbb{P}(Y'_d \in A)$ in pogoj sledi iz predpostavke $1 - pm > p$.
(4) $d \in A, d' \notin A$: Velja:

$$\mathbb{P}(Y_d \in A) = p(|A| - m - 1) + 1,$$
$$\mathbb{P}(Y'_d \in A) = p|A|.$$

Sedaj z uporabo potrebnega pogoja dobimo (hkrati upoštevamo, da je $|A| \geq 1$)

$$1 - pm \leq pe^{\epsilon} + \delta = e^{\epsilon}(p|A| - p|A| + p) + \delta \leq e^{\epsilon}p|A| - p(|A| - 1) + \delta,$$

kar po preureditvi da

$$p(|A| - m - 1) + 1 \leq e^{\epsilon}p|A| + \delta.$$

Torej vidimo, da bo naš mehanizem (ϵ, δ)-diferencirano zaseben natanko tedaj, ko bo $p \geq \frac{1-\delta}{m+e^{\epsilon}}$.

♢

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 eno-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 5.1. *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 5.2. 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 5.3. *Naj bo podana družina 1-dimenzionalnih diferencirano zasebnih mehanizmov $\{Y_d : \Omega \to U | d \in D\}$ (glej definicijo 1). Potem velja*

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

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

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 5.5. *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*

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

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 5.6. 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.

Primer 5.7. Vrnimo se na primer 4.10, kjer D predstavlja množico hobijev (diskreten primer podatkov, $|D| = m+1$). Pokazali smo, da obstaja (ϵ, δ) diferencirano zaseben mehanizem s $p = \frac{1-\delta}{m+e^{\epsilon}}$. Če je D opremljen z diskretno metriko, potem velja $\rho(d, d') = 1$ za vsak $d \neq d'$ in $\kappa = 1$. Potem za vsak $d \in D$ sledi:

$$\gamma = \mathbb{E}[\rho(Y_d, d)] = \sum_{d \neq d'} p = mp = (1 - \delta)\frac{m}{m + e^{\epsilon}}.$$

Vidimo, da je v tem primeru ocena za spodnjo mejo napake, ki jo dobimo iz izreka 5.5 optimalna. ♢

6. Dodatno o funkcijskih podatkih

Dosedaj smo že navedli primera mehanizmov za diskretne in numerične podatke (glej primera 4.9 ter 4.10). V tem poglavju bomo opisali mehanizem še za funkcijske podatke. Pri tem se bomo sklicali na v uvodu omenjeno teorijo Gaussovih procesov ter reprodukcijskih jeder Hilbertovih prostorov.

Obstajata dve glavni motivaciji za obravnavo funckijskih podatkov v okviru diferencirane zasebnosti. Prvo smo že omenili in sicer to, da je naša podatkovna baza D sestavljena iz funkcij (primer 3.3). Druga nastopi, ko želimo kot odgovor na poizvedbo Q podati funkcijo (torej je E_Q funkcijski prostor). Tak primer bi bil, ko imamo numerične podatke $d_1, ..., d_n \in \mathbb{R}^d$, ki so dobljeni iz porazdelitve z gostoto f, naš cilj pa je oceniti gostoto s ti. jedrno cenilko za gostoto (angl. kernel density estimator)

$$\hat{f}(x) = \frac{1}{n}\sum_{1}^{n} W\Big(\frac{||x - d_i||}{h}\Big), x \in \mathbb{R}^d,$$

kjer je W jedrna funkcija in h njen parameter dosega (angl. bandwith parameter). Na kratko, gre za parameter, ki vpliva na to, kako blizu si morata biti dani točki, da imata znaten vpliv druga na drugo. Oblika $\hat{f}$ je močno odvisna od izbire tega parametra in obstajajo načini za optimalen izbor, ki pa se jih v tem delu ne bomo dotaknili. S pomočjo mehanizma, ki ga bomo podali v nadaljevanju, lahko potem cenilko gostote $\hat{f}$ naredimo diferencirano zasebno. Koristnost takega pristopa je večplastna. Ne samo, da smo dobili zasebno cenilko za gostoto, ampak lahko sedaj le-to uporabimo tudi za vzorčenje novih podatkov. To je alternativen pristop k objavi celotne podatkovne baze kot tisti, ki smo ga obravnali pri primeru Laplacovega mehanizma za numerične podatke (primer 4.9). Na tako privatizirani bazi lahko potem izvajamo vse poizvedbe in pogojem diferencirane zasebnosti bo še vedno zadoščeno. Primer, ko je odgovor na statistično poizvedbo oz. analizo funkcija, so tudi razni modeli iz strojnega učenja (linearna regresija, metoda podpornih vektorjev itd.). Opisan postopek v tem poglavju se lahko uporabi tudi za ohranitev zasebnosti pri učenju takih modelov. Na ta način preprečimo, da bi prek naučenih parametrov modela izdali preveč podrobnosti o učni podatkovni množici.

Podan mehanizem ne bo pokril prve motivacije pri funkcijsih podatkih, torej primera, ko je naša podatkovna baza D sestavljena iz funkcij. Pri pregledu literature s področja diferencirane zasebnosti takih mehanizmov nisem zasledil.

Podajmo sedaj okvir, znotraj katerega bomo navedli nekaj osnovnih rezultatov. Naši podatki naj bodo numerični, torej $D = \mathbb{R}^m$. Prostor možnih odgovor je funkcijski in sicer se omejimo na funkcije, ki zavzemajo realne vrednosti $E_Q = \{f | f : \mathbb{R}^m \to \mathbb{R}\} \subset \mathbb{R}^D$. Odgovor na poizvedbo $Q : D^n \to E_Q$ označimo z $Q(\mathbf{d}) = f_{\mathbf{d}}$, kjer $\mathbf{d} = (d_1, ..., d_n) \in D^n$ kot prej predstavlja podatkovno bazo. Odzivni mehanizem ima obliko $X_{Q,\mathbf{d}} = X_{Q(\mathbf{d})} = \widetilde{f}_{\mathbf{d}}$, torej vidimo, da gre za perturbacijo odgovorov na poizvedbo (3). V naslednjem izreku podamo mehanizem, kjer diferencirano zasebnost dosežemo z uporabo Gaussovih procesov.

Izrek 6.1. *Naj bo G trajektorija Gaussovega procesa s povprečjem 0 in kovariančno funkcijo K. Naj bodo $x_1, ...x_n \in D$. Označimo z M Grammovo matriko*

$$M(x_1, ..., x_n) = \begin{pmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{pmatrix}.$$

Potem bo odzivni mehanizem

$$\widetilde{f_{\boldsymbol{d}}} = f_{\boldsymbol{d}} + \sqrt{2 \log \frac{2}{\delta}} \frac{\Delta}{\epsilon} G$$

(ϵ, δ) diferencirano zaseben, ko bo veljajo

$$\sup_{\boldsymbol{d} \sim \boldsymbol{d'}} \sup_{n < \infty} \sup_{(x_1, ..., x_n) \in D^n} \left\| M^{-1/2}(x_1, ...x_n) \begin{pmatrix} f_{\boldsymbol{d}}(x_1) - f_{\boldsymbol{d'}}(x_1) \\ \vdots \\ f_{\boldsymbol{d}}(x_n) - f_{\boldsymbol{d'}}(x_n) \end{pmatrix} \right\|_2 \leq \Delta. \tag{6}$$

Preden dokažemo zgornji izrek, si poglejmo še dve trditvi, ki predstavljata glavni ideji za dokazom. Izrek 6.2 obravnava Gaussov mehanizem v primeru, ko je odgovor na poizvedbo končno dimenzionalen numeričen vektor. Tak primer je npr. ko so elementi naše podatkovne baze numerični vektorji $d_i \in \mathbb{R}^d$, odgovor na poizvedbo pa je vektor povprečnih komponent $Q(\mathbf{d}) = \frac{1}{n} \sum_{i=1}^{n} d_i \in \mathbb{R}^d$.

Izrek 6.2. *Naj bo podana poizvedba $Q : \mathbb{R}^d \to \mathbb{R}^d$. Naj bo $M \in \mathbb{R}^{d \times d}$ pozitivno definitna simetrična matrika. Dalje naj za poizvedbo Q velja*

$$\sup_{\boldsymbol{d} \sim \boldsymbol{d'}} \|M^{-1/2}(Q(\boldsymbol{d}) - Q(\boldsymbol{d'})\|_2 \leq \Delta. \tag{7}$$

Dodatno predpostavimo še $\epsilon \leq 1$. Potem je odzivni mehanizem

$$X_{Q,\boldsymbol{d}} = X_{Q(\boldsymbol{d})} = Q(\boldsymbol{d}) + \sqrt{2 \log \frac{2}{\delta}} \frac{\Delta}{\epsilon} z, z \sim \mathcal{N}_d(0, M)$$

(ϵ, δ) diferencirano zaseben.

Za dokaz izreka 6.2 potrebujemo naslednjo lemo, ki pove, da je za dosego (ϵ, δ) diferencirane zasebnosti dovolj, da dosežemo $(\epsilon, 0)$ zasebnost na dovolj veliki množici (v smislu verjetnostne mere $\mathbb{P}$).

Lema 6.3. *Naj za vse sosednje podatkovne baze $\boldsymbol{d} \sim \boldsymbol{d'}$ obstaja množica $A^{\star}_{\boldsymbol{d},\boldsymbol{d'}} \in \mathcal{A}_Q$, za katero velja*

$$S \subseteq A^{\star}_{\boldsymbol{d},\boldsymbol{d'}} \Rightarrow \mathbb{P}(X_{Q,\boldsymbol{d}} \in S) \leq e^{\epsilon} \mathbb{P}(X_{Q,\boldsymbol{d'}} \in S) \tag{8}$$

in

$$\mathbb{P}(X_{Q,\boldsymbol{d}} \in A^{\star}_{\boldsymbol{d},\boldsymbol{d'}}) \geq 1 - \delta. \tag{9}$$

Potem je mehanizem $X_{Q,\boldsymbol{d}}$ (ϵ, δ) diferencirano zaseben.

Dokaz. Vzemimo poljubno množico $S \in \mathcal{A}_Q$. Sledi

$$\begin{aligned} \mathbb{P}(X_{Q,\mathbf{d}} \in S) &= \mathbb{P}(X_{Q,\mathbf{d}} \in (S \bigcap A^{\star}_{\mathbf{d},\mathbf{d'}})) + \mathbb{P}(X_{Q,\mathbf{d}} \in (S \bigcap A^{\star C}_{\mathbf{d},\mathbf{d'}})) \\ &\leq \mathbb{P}(X_{Q,\mathbf{d}} \in (S \bigcap A^{\star}_{\mathbf{d},\mathbf{d'}})) + \delta \leq e^{\epsilon} \mathbb{P}(X_{Q,\mathbf{d'}} \in (S \bigcap A^{\star}_{\mathbf{d},\mathbf{d'}})) + \delta \leq \\ &e^{\epsilon} \mathbb{P}(X_{Q,\mathbf{d}} \in S) + \delta. \end{aligned}$$

Prva neenakost sledi iz (9), druga iz (8) in tretja iz monotonosti verjetnostne mere. □

Opomba 6.4. Če je $X_{Q,\mathbf{d}}$ absolutno zvezna slučajna spremenljivka, lahko govorimo o njeni gostoti (v Radon-Nikodymjevem smislu) z ozirom na Lebesgueovo mero λ. Označimo jo z $g_{\mathbf{d}} = \frac{dF_{X_{Q,\mathbf{d}}}}{d\lambda}$, kjer je $F_{X_{Q,\mathbf{d}}}$ porazdelitveni zakon (v splošnem ni nujno da vzamemo Lebesguovo mero λ, zadostuje že, da vzamemo σ-končno mero, glede na katero je mera $dF_{X_{Q,\mathbf{d}}}$ absolutno zvezna). V tem primeru je zadosten pogoj za (8) to, da je razmerje gostot omejeno na dani množici $A^{\star}_{\mathbf{d},\mathbf{d'}}$:

$$\forall a \in A^{\star}_{\mathbf{d},\mathbf{d'}} : g_{\mathbf{d}}(a) \leq e^{\epsilon} g_{\mathbf{d'}}(a).$$

To sledi iz Radon-Nikodymjevega izreka in monotonosti integrala:

$$\mathbb{P}(X_{Q,\mathbf{d}} \in S) = \int_S g_{\mathbf{d}}(a) d\lambda(a) \leq \int_S e^{\epsilon} g_{\mathbf{d'}}(a) d\lambda(a) = e^{\epsilon}\mathbb{P}(X_{Q,\mathbf{d'}} \in S).$$

Dokaz izreka 6.2. Za večjo preglednost označimo v izreku definiran mehanizem $X_{Q,\mathbf{d}}$ z $\tilde{Q}(\mathbf{d})$ in $\sqrt{2\log\frac{2}{\delta}}$ s $c(\delta)$. Ker je z multivariatno normalno porazdeljena slučajna spremenljivka, je absolutno zvezna. Posledično enako velja za slučajno spremenljivko/odzivni mehanizem $\tilde{Q}(\mathbf{d})$ in torej imamo gostoto. Sledi

$$\frac{dF_{\tilde{Q}(\mathbf{d})}(x)/d\lambda(x)}{dF_{\tilde{Q}(\mathbf{d'})}(x)/d\lambda(x)} = \frac{g_{\mathbf{d}(x)}}{g_{\mathbf{d'}(x)}}$$

$$= exp\Big\{ \frac{\epsilon^2}{2c(\delta)^2\Delta^2} \Big[(x - Q(\mathbf{d'}))^T M^{-1}(x - Q(\mathbf{d'})) - (x - Q(\mathbf{d}))^T M^{-1}(x - Q(\mathbf{d})) \Big] \Big\}.$$

Razmerje gostot bo večje od e^{ϵ} le v primeru

$$2x^T M^{-1}(Q(\mathbf{d}) - Q(\mathbf{d'})) + Q(\mathbf{d'})^T M^{-1} Q(\mathbf{d'}) - Q(\mathbf{d})^T M^{-1} Q(\mathbf{d}) \geq 2\frac{c(\delta)^2\Delta^2}{\epsilon}.$$

Pokažimo sedaj, da je verjetnost (oz. mera) te množice manjša od δ, iz česar bo nato po lemi 6.3 sledila (ϵ, δ) diferencirana zasebnost $\tilde{Q}(\mathbf{d})$ in izrek bo dokazan.

Ker delamo z mehanizmom oblike $\tilde{Q}(\mathbf{d})$ vzamemo $x = Q(\mathbf{d}) + c(\delta)\frac{\Delta}{\epsilon} M^{-1/2} u, u \sim \mathcal{N}_d(0, I)$. Dodatno še množimo obe strani neenakosti z $\frac{\epsilon}{c(\delta)\Delta}$ in upoštevamo (7), kar nam da

$$u^T M^{-1/2}(Q(\mathbf{d}) - Q(\mathbf{d'})) \geq \frac{c(\delta)\Delta}{\epsilon} - \frac{\epsilon\Delta}{2c(\delta)}.$$

Opazimo, da na levi strani neenakosti dobimo normalno porazdeljeno slučajno spremenljivko s povprečjem 0 in varianco manjšo od Δ^2 (sledi iz (7)). Označimo $y \sim \mathcal{N}(0, 1)$. Sledi

$$\mathbb{P}\Big(u^T M^{-1/2}(Q(\mathbf{d}) - Q(\mathbf{d'})) \geq \frac{c(\delta)\Delta}{\epsilon} - \frac{\epsilon\Delta}{2c(\delta)}\Big) \leq$$

$$\mathbb{P}\Big(\Delta y \geq \frac{c(\delta)\Delta}{\epsilon} - \frac{\epsilon\Delta}{2c(\delta)}\Big) \leq \mathbb{P}\Big(y \geq c(\delta) - \frac{1}{2c(\delta)}\Big) \leq \delta.$$

Prva neenakost sledi iz dejstva, da se verjetnost dogodka povečuje z varianco δ, druga iz predpostavke $\epsilon \leq 1$ in zadnja iz neenakosti za repe porazdelitve standarne normalne slučajne spremenljivke $\mathbb{P}(Y \geq y) \leq \frac{\exp(-y^2/2)}{y\sqrt{2\pi}}$ (za dokaz glej [8]). □

Dotaknimo se še σ-algebre v primeru, ko je prostor odgovorov na poizvedbe funkcijski, torej $E_Q = \{f | f : D \to \mathbb{R}\} = \mathbb{R}^D$ (pri čemer je $D = \mathbb{R}^m$). Definirajmo najprej ti. cilindrične množice funkcij

$$C_{S,B} = \{f \in \mathbb{R}^D : (f(x_1), ..., f(x_n)) \in B\},$$

kjer je $S = (x_1, ..., x_n)$ končen nabor točk iz D in $B \in \mathcal{B}(\mathbb{R}^n)$. Vidimo, da gre za množice funkcij, ki če jih izvrednotimo v točkah iz S, zavzamejo vrednosti v Borelovi množici B. Za izbrani S označimo $\mathcal{C}_S = \{C_{S,B} : B \in \mathcal{B}(\mathbb{R}^n)\}$. Unija po vseh končnih množicah S

$$\bigcup_{S:|S|<\infty} \mathcal{C}_S =: \mathcal{F}_0$$

je algebra (glej stran 485 v [10]), ni pa σ-algebra, saj nimamo zaprtosti za števne unije. Označimo še $\mathcal{F} = \sigma(\mathcal{F}_0)$. Podajmo sedaj izrek, ki pokaže, da je dovolj, da pogoj diferencirane zasebnosti preverimo le na algebri $\mathcal{F}_0$.

Izrek 6.5. *Naj bo $X_{Q,\boldsymbol{d}}$ odzivni mehanizem z $E_Q = \mathbb{R}^D$ (ponovno $D = \mathbb{R}^m$) in $\mathcal{A}_Q = \mathcal{F}$. Če velja pogoj diferencirane zasebnosti*

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

za vsaki $\boldsymbol{d} \sim \boldsymbol{d'}$ in za vsako $A \in \mathcal{F}_0$, potem je dan mehanizem (ϵ, δ) diferencirano zaseben (torej velja pogoj za vsak $A \in \mathcal{F}$).

Dokaz. Za števno množico S cilindrična σ-algebra zavzame obliko

$$C_{S,B} = \{f \in \mathbb{R}^D : f(x_i) \in B_i, i = 1, 2, ...\} = \bigcap_{i=1}^{\infty} C_{\{x_i\},B_i},$$

kjer so B_i Borelove množice na $\mathbb{R}$. Definirajmo $C_{S,B,n} = \bigcap_{i=1}^{n} C_{\{x_i\},B_i}$ in $C_{S,B} = \bigcap_{n=1}^{\infty} C_{S,B,n}$. Ker je zaporedje množic $C_{S,B,n}$ padajoče iz zveznosti verjetnosti mere sledi

$$\mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B}) = \mathbb{P}(X_{Q,\mathbf{d}} \in \bigcap_{n=1}^{\infty} C_{S,B,n}) = \lim_{n\to\infty} \mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B,n}).$$

Posledično za vsaki sosednji bazi $\mathbf{d} \sim \mathbf{d'}$ in za vsak $\alpha > 0$ obstaja n_0, tako da za vsak $n \geq n_0$ velja

$$|\mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B}) - \mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B,n})| \leq \alpha,$$
$$|\mathbb{P}(X_{Q,\mathbf{d'}} \in C_{S,B}) - \mathbb{P}(X_{Q,\mathbf{d'}} \in C_{S,B,n})| \leq \alpha.$$

Nadalje dobimo

$$\begin{aligned}\mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B}) &\leq \mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B,n_0}) + \alpha \\ &\leq e^\epsilon \mathbb{P}(X_{Q,\mathbf{d'}} \in C_{S,B,n_0}) + \alpha + \delta \\ &\leq e^\epsilon \mathbb{P}(X_{Q,\mathbf{d'}} \in C_{S,B}) + \alpha(1 + e^\epsilon) + \delta \\ &\leq e^\epsilon \mathbb{P}(X_{Q,\mathbf{d'}} \in C_{S,B}) + b\alpha + \delta,\end{aligned}$$

pri čemer smo privzeli $(1 + e^\epsilon) < b$. V kontekstu diferencirane zasebnosti je ϵ navadno majhen (med 0 in 1), zato ta privzetek ni omejujoč. Ker zgornje velja za vsak $\alpha > 0$, dobimo $\mathbb{P}(X_{Q,\mathbf{d}} \in C_{S,B}) \leq e^\epsilon \mathbb{P}(X_{Q,\mathbf{d'}} \in C_{S,B}) + \delta$ in izrek sledi. □

Opomba 6.6. Izrek 6.5 bi lahko dokazali tudi drugače in sicer z uporabo izreka 4.1 o zadostnih testnih množicah. Velja namreč $\mathcal{F} = \sigma(\mathcal{F}_0)$.

Opomba 6.7. Če je $E_Q = C([0,1])$ in za normo vzamemo $\|f\|_\infty = sup\{|f(x)| : x \in [0,1]\}$ se zgoraj omenjena teorija o cilindrični σ-algebri $\mathcal{F}$ navezuje na primer 4.2.

Sedaj smo pripravili vse potrebno in lahko dokažemo uvodni izrek v tem poglavju.

Dokaz izreka 6.1. Dokazali bomo, da je podan mehanizem $X_{Q,\mathbf{d}} = \tilde{f}_{\mathbf{d}}$ diferencirano zaseben glede na $\mathcal{A}_Q = \mathcal{F}$, torej cilindrično σ-algebro. Za končen nabor točk $(x_1, ..., x_n) \in D^n$ je vektor $(G(x_1), ..., G(x_2))$ porazdeljen multivariatno normalno s povprečjem 0 in kovariančno matriko enako M (lastnost Gaussovih procesov). Torej za končno dimenzionalen vektor, ki ga dobimo, ko mehanizem oz. funkcijo $\tilde{f}_{\mathbf{d}}$ izvrednotimo v točkah $(x_1, ..., x_n)$, nam diferencirano zasebnost zagotavlja izrek 6.2, saj pogoj (6) implira omejitev občutljivosti poizvedbe (7). Torej za vsak $n < \infty$, $(x_1, ..., x_n) \in D^n$, $B \in \mathcal{B}(\mathbb{R}^n)$ velja

$$\mathbb{P}((\tilde{f}_{\mathbf{d}}(x_1), ..., \tilde{f}_{\mathbf{d}}(x_n)) \in B) \leq e^{\epsilon}\mathbb{P}((\tilde{f}_{\mathbf{d'}}(x_1), ..., \tilde{f}_{\mathbf{d'}}(x_n)) \in B) + \delta,$$

za vsaki $\mathbf{d} \sim \mathbf{d'}$. Dalje opazimo, da lahko vsako množico $A \in \mathcal{F}_0$ zapišemo kot $A = C_{X_n,B}$ za nek končnen n, $(x_1, ..., x_n) \in D^n$ in Borelovo množico B. Potem

$$\mathbb{P}(X_{Q,\mathbf{d}} \in A) = \mathbb{P}(\tilde{f}_{\mathbf{d}} \in A) = \mathbb{P}((\tilde{f}_{\mathbf{d}}(x_1), ..., \tilde{f}_{\mathbf{d}}(x_n)) \in B).$$

S tem smo pokazali, da pogoj velja za vse $A \in \mathcal{F}_0$. Izrek 6.5 razširi diferencirano zasebnost na celotno $\mathcal{F}$. □

Izkaže se, da se da pogoju (6) iz izreka 6.1 enostavno zadostiti, če funkcije $f_{\mathbf{d}}$ ležijo v Hilbertovem prostoru z reprodukcijskim jedrom, ki je enak kovariančni funkciji Gaussovega procesa. To pokažeta naslednji izrek in pomembna posledica.

Izrek 6.8. *Naj bo $f \in \mathcal{H}$, kjer je $\mathcal{H}$ Hilbertov prostor z reprodukcijskim jedrom K. Potem za vsako $x_1, ..., x_n$ končno zaporedje točk v $\mathbb{R}^m$ velja:*

$$\left\| M^{-1/2}(x_1, ...x_n) \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} \right\|_2 \leq \|f\|_{\mathcal{H}}.$$

Dokaz. Definirajmo operator $P : \mathcal{H} \to \mathcal{H}$ kot

$$P = \sum_{i=1}^{n} K_{x_i} \sum_{j=1}^{n} (M^{-1})_{i,j} \langle K_{x_j}, \cdot \rangle_{\mathcal{H}}$$

Najprej opazimo, da je ta operator idempotenten, torej $P = P^2$:

$$\begin{aligned}
P^2 &= \sum_{i=1}^{n} K_{x_i} \sum_{j=1}^{n} (M^{-1})_{i,j} \langle K_{x_j}, \sum_{k=1}^{n} K_{x_k} \sum_{l=1}^{n} (M^{-1})_{k,l} \langle K_{x_l}, \cdot \rangle_{\mathcal{H}} \rangle_{\mathcal{H}} \\
&= \sum_{i=1}^{n} K_{x_i} \sum_{j=1}^{n} (M^{-1})_{i,j} \sum_{k=1}^{n} \langle K_{x_j}, K_{x_k} \rangle_{\mathcal{H}} \sum_{l=1}^{n} (M^{-1})_{k,l} \langle K_{x_l}, \cdot \rangle_{\mathcal{H}} \\
&= \sum_{i=1}^{n} K_{x_i} \sum_{j=1}^{n} (M^{-1})_{i,j} \sum_{k=1}^{n} M_{j,k} \sum_{l=1}^{n} (M^{-1})_{k,l} \langle K_{x_l}, \cdot \rangle_{\mathcal{H}} \\
&= \sum_{i=1}^{n} K_{x_i} \sum_{l=1}^{n} (M^{-1})_{i,l} \langle K_{x_l}, \cdot \rangle_{\mathcal{H}} = P
\end{aligned}$$

Pri drugem enačaju smo uporabili linearnost skalarnega produkta, pri tretjem pa definicijo Grammove matrike. Zlahka tudi preverimo, da je P sebi-adjungiran operator $\langle Pf, g \rangle_{\mathcal{H}} = \langle f, Pg \rangle_{\mathcal{H}}$ (posledica simetrije matrike M). Dalje sledi

$$\begin{aligned}
\|f\|_{\mathcal{H}}^2 &= \langle f, f \rangle_{\mathcal{H}} = \langle Pf + (f - Pf), Pf + (f - Pf) \rangle_{\mathcal{H}} \\
&= \langle Pf, Pf \rangle_{\mathcal{H}} + 2\langle Pf, f - Pf \rangle_{\mathcal{H}} + \langle f - Pf, f - Pf \rangle_{\mathcal{H}} \\
&= \langle Pf, Pf \rangle_{\mathcal{H}} + 2\langle f, Pf - P^2 f \rangle_{\mathcal{H}} + \langle f - Pf, f - Pf \rangle_{\mathcal{H}} \\
&= \langle Pf, Pf \rangle_{\mathcal{H}} + \langle f - Pf, f - Pf \rangle_{\mathcal{H}} \\
&\geq \langle Pf, Pf \rangle_{\mathcal{H}} = \langle f, Pf \rangle_{\mathcal{H}}
\end{aligned}$$

□

Posledica 6.9. *Naj E_Q (prostor možnih odgovorov, torej funkcij, na poizvedbo) podmnožica Hilbertovega prostora $\mathcal{H}$ z reprodukcijskim jedrom K, ki je enak kovariančni funkciji Gaussovega procesa. Če z G označimo trajektorijo tega Gaussovega procesa, potem bo mehanizem*

$$\widetilde{f}_{\boldsymbol{d}} = f_{\boldsymbol{d}} + \sqrt{2 \log \frac{2}{\delta}} \frac{\Delta}{\epsilon} G$$

(ϵ, δ) diferencirano zaseben, ko bo veljajo

$$\sup_{\boldsymbol{d} \sim \boldsymbol{d'}} \|f_{\boldsymbol{d}} - f_{\boldsymbol{d'}}\|_{\mathcal{H}} \leq \Delta.$$

Primer 6.10. Uporabimo zgoraj omenjeni mehanizem na konkretnem primeru. Recimo, da naši podatki $(d_1, ..., d_n) = \mathbf{d}, d_i \in \mathbb{R}^d$ prihajajo iz porazdelitve z gostoto f. Za oceno te gostote uporabimo cenilko z Gaussovim jedrom

$$f_{\mathbf{d}}(x) = \frac{1}{n(2\pi h^2)^{d/2}} \sum_{i=1}^{n} exp\Big\{ \frac{-\|x - d_i\|_2^2}{2h^2} \Big\}, x \in \mathbb{R}^d,$$

kjer je h parameter dosega. Naj bosta $\mathbf{d} \sim \mathbf{d'}$ sosednji bazi, ki se razlikujeta le na zadnjem elementu (brez izgube za splošnost), torej $\mathbf{d'} = (d_1, ..., d_{n-1}, d'_n)$. Sledi

$$(f_{\mathbf{d}} - f_{\mathbf{d'}})(x) = \frac{1}{n(2\pi h^2)^{d/2}} \Big(exp\Big\{ \frac{-\|x - d_n\|_2^2}{2h^2} \Big\} - exp\Big\{ \frac{-\|x - d'_n\|_2^2}{2h^2} \Big\} \Big).$$

Za kovariančno funkcijo Gaussovega procesa $\{X_t : t \in T\}$ vzemimo Gaussovo jedro $K(x, y) = exp\{\frac{-\|x-y\|_2^2}{2h^2}\}$. Hkrati opazimo, da funkcije oblike $f_{\mathbf{d}}$ ležijo v Hilbertovem

prostoru z reprodukcijskim jedrom K. Torej $f_{\mathbf{d}} - f_{\mathbf{d'}} = \frac{1}{n(2\pi h^2)^{d/2}}(K_{d_n} - K_{d'_n})$ in velja

$$\begin{aligned}\|f_{\mathbf{d}} - f_{\mathbf{d'}}\|^2_{\mathcal{H}} &= \langle f_{\mathbf{d}} - f_{\mathbf{d'}}, f_{\mathbf{d}} - f_{\mathbf{d'}}\rangle_{\mathcal{H}} \\ &= \Big(\frac{1}{n(2\pi h^2)^{d/2}}\Big)^2 \langle K_{d_n} - K_{d'_n}, K_{d_n} - K_{d'_n}\rangle_{\mathcal{H}} \\ &= \Big(\frac{1}{n(2\pi h^2)^{d/2}}\Big)^2 (K(d_n, d_n) + K(d'_n, d'_n) - 2K(d_n, d'_n)) \\ &\leq 2\Big(\frac{1}{n(2\pi h^2)^{d/2}}\Big)^2\end{aligned}$$

Če sedaj z G označimo trajektorijo Gaussovega procesa s povprečjem 0 in kovariančno funkcijo enako Gaussovemu jedru, potem iz posledice 6.9 sledi, da je mehanizem

$$\widetilde{f}_{\mathbf{d}} = f_{\mathbf{d}} + \sqrt{\log\frac{2}{\delta}}\frac{1}{\epsilon n(2\pi h^2)^{(d/2)}}G$$

(ϵ, δ) diferencirano zaseben. ♢

7. Komentar praktičnega dela

Za konec si poglejmo še, kako se do zdaj omenjeni mehanizmi obnesejo v praksi. V ta namen bomo uporabili umetno zgenerirane podatke o 1000 prebivalcih ZDA, ki vsebujejo informacije o njihovi plači, telesni višini in zvezni državi, v kateri živijo. Vsa koda s komentarji je dostopna na naslovu: `https://github.com/metodj/thesis/blob/master/dfPython.ipynb` .

7.1. **Numerični podatki.** Najprej uporabimo Laplacov mehanizem iz primera 4.9 za podatke o plači. Višina plača je bila dobljena iz vzorčenja enakomerne porazdelitve na intervalu [1500\$, 4500\$]. Za metriko vzamemo absolutno razdaljo. Diameter naših podatkov, ki je v tem primeru kar enak razliki maksimalne in minimalne vrednosti plače v bazi, potem znaša 2996\$. Rezultati so prikazani v spodnjih dveh tabelah.

ϵ	δ	b	povprečna razlika
0.1	0.1	14589	14935
2	0.5	1112	1075
11	0.5	245	249

Tabela 2. b predstavlja parameter Laplacove porazdelitve, povprečna razlika pa povprečje odstopanj diferencirano zasebnih podatkov od prvotnih.

Vidimo, da moramo za dosego zasebnosti pri običajnih vrednostih parametrov ($\epsilon = 0.1, \delta = 0.1$) uporabiti zelo 'razpršeno' Laplacovo spremenljivko (visoka vrednost parametra b pomeni veliko varianco). Uporabnost podatkov se zaradi tega skoraj povsem izgubi, kar se vidi tudi iz tabele 3. Minimalna vrednost plače v podatkovni bazi tako znaša -111499\$, kar je seveda povsem nesmiselno. Približno smiselne rezultate dobimo pri zelo visokih vrednostih ϵ in δ. Če izvzamemo dejstvo, da $\delta = 0.7$ pomeni, da se bo naš mehanizem zlomil s kar 70% verjetnostjo, parameter $\epsilon = 11$ pove to, da se lahko verjetnosti iz definicije diferencirane zasebnosti

razlikujeta za faktor približno 60000. To seveda pomeni, da čeprav se natančnost zdi približno smiselna, je tak mehanizem za zaščito zasebnosti neuporaben. Na tem

	povprečje	min	max
prvotni podatki	2995	1504	4500
(0.1 , 0.1)	3402	-111499	109729
(2, 0.5)	2999	-6110	11411
(11, 0.7)	2990	765	5360

TABELA 3. Vrednosti nekaterih osnovih poizvedb pri različnih vrednostih parametrov (ϵ, δ).

mestu lahko uporabimo še izrek 5.3, ki nam da spodnjo mejo za največjo napako obravnavanega mehanizma. Rezultati se nahajajo v tabeli 4. Opazimo, da je sicer res, da so dejanske največje napake večje od spodnjih mejih, ki nam jih da izrek, vendar so te meje precej neoptimalne in posledično neinformativne.

(ϵ, δ)	(0.1, 0.1)	(2, 0.5)	(11, 0.7)
γ	640	89	0.0075
dejanska največja napaka	110600	7203	1803

TABELA 4. Spodnja meja največje napake γ in dejansko opažena največja napaka pri različnih vrednostih parametrov (ϵ, δ).

Opomba 7.1. Razlog za slabe rezultate zgoraj obravnavanega mehanizma leži v veliki občutljivosti identične poizvedbe. Občutljivost (angl. sensitivity) nam pove, kako močno se odgovor na dano poizvedbo spremeni glede na prisotnost konkretnega posameznika v podatkovni bazi. Matematično se definira na različne načine, v primeru numeričnih podatkov ($D = \mathbb{R}, E_Q = \mathbb{R}$) pa se pogosto definira kot

$$\Delta Q = \max_{\mathbf{d} \sim \mathbf{d'}} \|Q(\mathbf{d}) - Q(\mathbf{d'})\|_1,$$

kjer $\|\cdot\|_1$ predsavlja l_1 normo. V primeru identične poizvedbe je občutljivost enaka kar $diam(D)$. Logično je, da gre za najbolj občutljivo poizvebo, saj 'izdamo' največ informacij o naši podatkovni bazi in posledično moramo dodati veliko šuma za dosego diferencirane zasebnosti. V praksi se zato tak pristop uporablja le redko.

Opomba 7.2. Dober primer večje implementacije diferencirane zasebnosti v praksi je sistem za zaščito SQL-poizvedb, ki so ga razvili pri podjetju Uber [4]. Sistem za vsako poizvedbo (SQL-stavek) najprej izračuna občutljivost in glede na to doda potem kalibriran šum odgovoru na poizvedbo (gre torej za mehanizem oblike (3)). Na ta način so dosegli dobro razmerje med uporabnostjo in zasebnostjo mehanizmov.

7.2. **Diskretni podatki.** Uporabimo mehanizem iz primera 4.10 na podatkih o zveznih državah, v katerih živijo posamezniki v naši podatkovni bazi. Rezultati so prikazani v tabelah 5 in 6. Spodnjo mejo za največjo napako nam tu poda izrek 5.5. Vidimo, da so ocene za spodnjo mejo precej bolj optimalne kot v primeru numeričnih podatkov (sledi iz opombe 5.7). Dodajmo še, da če bi delali na podatkih o spolu (možna odgovora na poizvedbo sta le dva, moški ali ženski spol), bi imel naš mehanizem preprosto obliko Bernoullijeve slučajne spremenljivke.

(ϵ, δ)	verjetnost, s katero podamo pravi odgovor
(0.1, 0.1)	0.12
(2, 0.5)	0.57
(7, 0.6)	0.98

TABELA 5. Rezultati mehanizma za diskretne podatke.

(ϵ, δ)			
prvotni podatki	TX (114)	CA (102)	NY (63)
(0.1, 0.1)	TX (35)	OH (31)	CA (30)
(2, 0.5)	TX (68)	CA (62)	FL (45)
(7, 0.6)	TX (113)	CA (102)	NY (62)

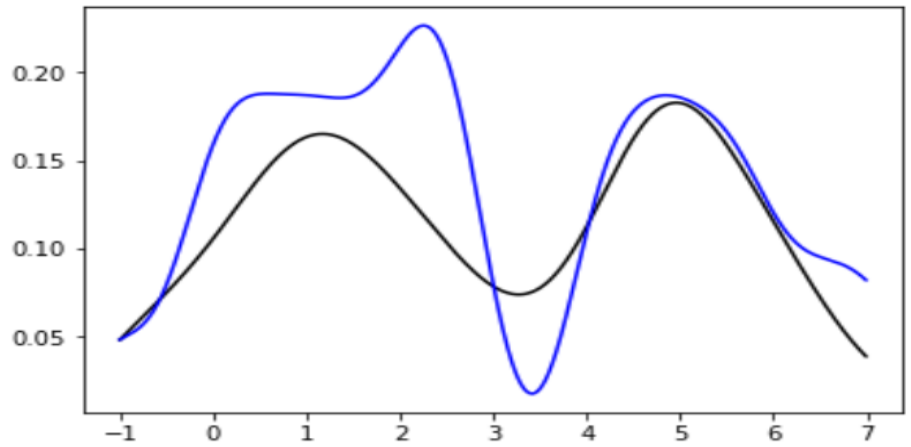
TABELA 6. Najpogostejše tri zvezne države v prvotnih podatkih in pri različnih vrednostih parametrov (ϵ, δ).

(ϵ, δ)	(0.1, 0.1)	(2, 0.5)	(7, 0.6)
γ	0.879	0.432	0.016
dejanska največja napaka	0.880	0.437	0.018

TABELA 7. Podatki o spodnjih mejah pri diskretnih podatkih. Dejanska največja napaka je tu izračunana kot razmerje med številom nepravilnih odgovorov ter številom posameznikov v bazi.

7.3. **Funkcijski podatki.** Na tem mestu smo implementirali mehanizem, ki vrne diferencirano zasebno jedrno cenilko za gostoto po vzoru primera 6.10. Rezultati so prikazani na slikah 1-3. Prvotni podatki so generirani iz mešanice dveh normalnih porazdelitev in sicer $\mathcal{N}(1,1)$ ter $\mathcal{N}(5,1)$. Pri vseh slikah je $\delta = 0.1$, spreminja se le parameter ϵ. Diferencirano zasebna cenilka za gostoto je prikazana z modro barvo.

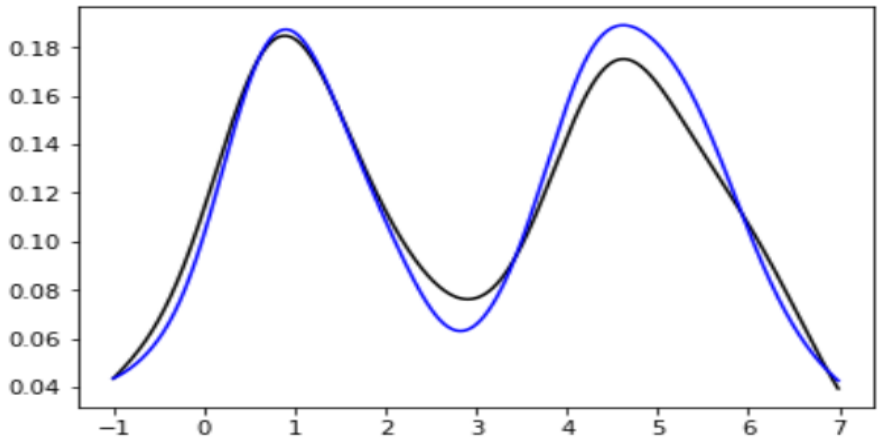
SLIKA 1. $\epsilon = 0.1$



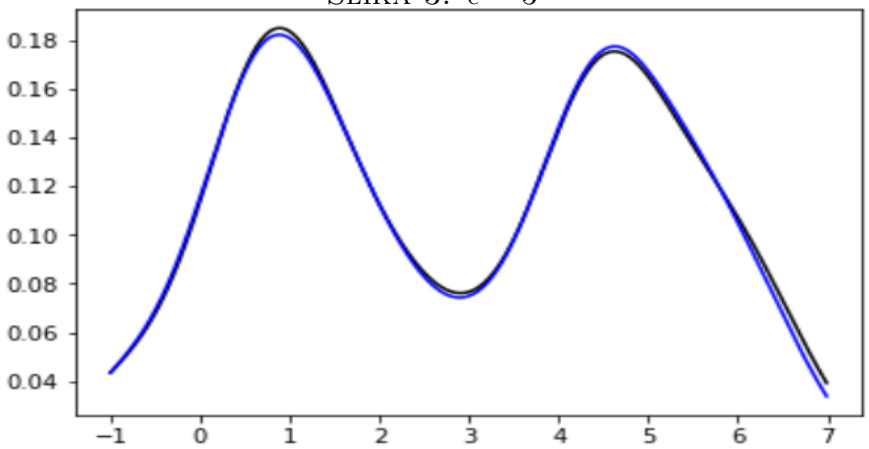
LITERATURA

[1] N. Holohan, D. J. Leith in O. Mason *Numerical, categorical and functional data under the one roof*, Information Sciences **305** (2015) strani 256–268.

Slika 2. $\epsilon = 1$



Slika 3. $\epsilon = 3$



[2] R. Hall, A. Rinaldo in L. Wasserman *Differential Privacy for Functions and Functional Data*, Journal of M. L. Research **14** (2013) strani 703–727.

[3] Differential Privacy Team, Apple *Learning with Privacy at Scale*, [ogled 22.07.2018], dostopno na `https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf`.

[4] N. Johnson, J. P. Near in D. Song *Towards Practical Differential Privacy for SQL Queries* , [ogled 22.07.2018], dostopno na `https://arxiv.org/abs/1706.09479`.

[5] *Reproducing kernel Hilbert space*, v: Wikipedia: The Free Encyclopedia, [ogled 22. 7. 2018], dostopno na `https://en.wikipedia.org/wiki/Reproducing_kernel_Hilbert_space`.

[6] *Differential Privacy*, v: Wikipedia: The Free Encyclopedia, [ogled 22. 7. 2018], dostopno na `https://en.wikipedia.org/wiki/Differential_Privacy`.

[7] *Gaussian Process*, v: Wikipedia: The Free Encyclopedia, [ogled 22. 7. 2018], dostopno na `https://en.wikipedia.org/wiki/Gaussian_process`.

[8] *Upper-tail inequality for standard normal distribution*, v: Mathematics Stack Exchange, [ogled 22. 7. 2018], dostopno na `https://math.stackexchange.com/questions/28751/proof-of-upper-tail-inequality-for-standard-normal-distribution/69417#69417`.

[9] K. Parthasarathy, *Probability Measures on Metric Spaces*, AMS Chelsea Publishing, 2005 (ponatis).

[10] P. Billingsley, *Probability and Measure*, Wiley–Interscience, 1995 (3. izdaja); dostopno tudi na naslovu `https://www.colorado.edu/amath/sites/default/files/attached-files/billingsley.pdf`.