UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Matematika – raziskovalna smer

Martin Raič

Chen–Steinova metoda

Magistrsko delo

Ljubljana, 1998

Kazalo

Program magistrskega dela

Aproksimacija porazdelitev slučajnih spremenljivk je eden od pomembnih problemov verjetnostnega računa. Chen–Steinova metoda je nov pristop k temu problemu, s katerim lahko po eni strani dokažemo klasične izreke o šibki konvergenci proti Poissonovi ali normalni porazdelitvi, po drugi strani pa v nasprotju z ostalimi metodami omogoča tudi, da ocenimo hitrost konvergence in kakovost aproksimacije.

V magistrskem delu prikažite idejo Chen–Steinove metode s pomočjo Dynkinove formule za markovske procese. Predstavite tudi primerno ozadje iz teorije operatorskih polgrup. Pri tem se oprite na vira [1] in [2]. Metodo nato uporabite pri Poissonovi aproksimaciji in pokažite, da je uporabna tudi za primer odvisnih slučajnih spremenljivk. Kot vodilo naj vam služijo viri [3], [4] in [5]. Delovanje Chen–Steinove metode prikažite tudi na primeru normalne aproksimacije tako za eno- kot tudi za večrazsežni primer. Oprite se na vira [3] in [6].

Literatura:

[1] L. C. G. Rogers, David Williams: *Diffusions, Markov Processes, and Martingales*, Volume 1: *Foundations*, John Willey & Sons, 1994.

[2] Stewart N. Ethier, Thomas G. Kurtz: *Markov Processes: Characterization and Convergence*, John Willey & Sons, 1986.

[3] A. D. Barbour: *Stein's Method*, neobjavljen rokopis, 1995. Dosegljiv na `ftp:iamassi.unizh.ch/pub/Barbour`.

[4] A. D. Barbour, Lars Holst, Svante Janson: *Poisson Approximation*, Clarendon Press, Oxford, 1992

[5] Richard Arratia, Larry Goldstein, Louis Gordon: *Poisson approximation and the Chen–Stein method*, Statistical Science **5** (1990), 403–434

[6] Charles Stein: *Approximate Computation of Expectations*, IMS, Hayward, Calif., 1986

V Ljubljani, 1. oktobra 1997

doc. dr. Mihael Perman

Povzetek

Delo obravnava Chen–Steinovo metodo, ki pomeni preokret pri aproksimaciji porazdelitev vsot slučajnih spremenljivk s Poissonovo, normalno ali tudi kako drugo "limitno" porazdelitvijo. Glavni prednosti te metode sta, da dá eksplicitno oceno napake in da ne deluje le za vsote neodvisnih, temveč tudi rahlo odvisnih slučajnih spremenljivk.

Ko porazdelitev dane slučajne spremenljivke X aproksimiramo z bolj znano porazdelitvijo μ, napako ocenjujemo z izrazi $|\mathbf{E}(f(X)) - \int f\, d\mu|$ za primeren razred testnih funkcij f. Osnovna ideja Chen–Steinove metode je, da poiščemo markovski (Feller–Dynkinov) proces, ki ima porazdelitev μ za ravnovesno, in rešimo Steinovo enačbo $Gg = f - \int f\, d\mu$, kjer je G generator operatorske polgrupe, ki pripada našemu markovskemu procesu. Tedaj velja $\mathbf{E}(f(X)) - \int f\, d\mu = \mathbf{E}(Gg)(X)$, torej je dovolj oceniti slednji izraz. Izkaže se, da je to bistveno lažje od ocene prvotnega izraza.

Delo predstavi aproksimacijo s Poissonovo in normalno porazdelitvijo, čeprav Chen–Steinova metoda nikakor ni omejena le na aproksimacijo s tema dvema porazdelitvama. Pri Poissonovi aproksimaciji navede tudi dva zgleda z odvisnimi slučajnimi porazdelitvami. Predstavi pa tudi markovske procese, ki tičijo v ozadju Chen–Steinove metode.

Math. Subj. Class (1991): 60F05, 60J35, 60J40

Key words: Stein's Method; Poisson Approximation; Central Limit Theorem; Markov Processes

Zahvala

Rad bi se zahvalil vsem, ki so kakor koli pripomogli k nastanku tega dela. V največji meri je to prav gotovo mentor, doc. dr. Mihael Perman, ki je s svojimi izvrstnimi podiplomskimi predavanji vnovič obudil moje zanimanje za verjetnost in se potrudil izbrati štiri obetavne teme za magistrsko delo, med katerimi je bila tudi Chen–Steinova metoda. Zahvaljujem se mu za branje tega dela in vse pripombe. Omogočil mi je tudi dostop do virov [4] in [24], brez katerih bi bilo to magistrsko delo precej revnejše. Tudi na koncu, ko se mi je že precej mudilo, mi je šel zelo na roke.

Hvala tudi prof. Legiši, ki je bil pripravljen uvrstiti predstavitev tega dela na Seminar iz funkcionalne analize, in ostalim udeležencem tega seminarja za vse pripombe. Navsezadnje pa gre zahvala tudi mojim dragim staršem, ki so mi dajali vso podporo.

Uvod

V teoriji verjetnosti (in seveda tudi v praksi) imamo precej opravka z vsotami slučajnih spremenljivk. S porazdelitvami vsot (denimo neodvisnih) slučajnih spremenljivk pa ni prav lahko računati, tudi če porazdelitve posameznih slučajnih spremenljivk dobro poznamo. Že na temeljno vprašanje, kolikšna je verjetnost, da vsota slučajnih spremenljivk pripada danemu intervalu, niti v najenostavnejših primerih ne moremo podati odgovora, ki bi dal lahko izračunljiv in eksakten numeričen rezultat. Zato se je seveda smiselno vprašati, ali obstaja vsaj kaka približna formula. Drugo vprašanje, ki se nam takoj porodi, pa je, kolikšno napako smo pri tem naredili.

Za vsoto neodvisnih in enako porazdeljenih Bernoullijevih slučajnih spremenljivk, torej za binomsko porazdelitev $\mathrm{Bin}(n,p)$, ki nam pove, koliko dogodkov se je zgodilo v n neodvisnih poskusih, pri čemer se vsak dogodek zgodi z verjetnostjo p, je odgovor že dolgo znan. Že de Moivre in Laplace sta izpeljala *lokalno* in *globalno* aproksimacijsko formulo: če je $X \sim \mathrm{Bin}(n,p)$, velja:

$$\mathbf{P}[X=k] = \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

kjer je $q = 1-p$, in:

$$\mathbf{P}[a \le X \le b] \approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{npq}}\right)$$

kjer je Φ znani verjetnostni integral:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt$$

Oznake za Φ niso enotne: morda je običajneje vzeti spodnjo mejo 0, nekateri pa vzamejo celo dvakrat večje vrednosti; mi se bomo ves čas držali zgornje formule.

Kar smo pravkar navedli, je v resnici aproksimacija z normalno porazdelitvijo: porazdelitev slučajne spremenljivke X smo aproksimirali z normalno porazdelitvijo $\mathrm{N}(np, npq)$. Pri fiksnem p je aproksimacija tem natančnejša, čim večji je n. Če je p majhen (npr. obratno sorazmeren n), pa je ugodneje uporabiti obrazec, pri katerem porazdelitev slučajne spremenljivke X aproksimiramo s Poissonovo porazdelitvijo $\mathrm{Po}(np)$ (to je tako imenovani *zakon majhnih števil*):

$$\mathbf{P}[X=k] \approx \frac{(np)^k e^{-np}}{k!}$$

Zadeva pa se takoj zaplete, če slučajne spremenljivke niso Bernoullijeve ali pa niso enako porazdeljene. Za poljubno, toda enako porazdeljene slučajne spremenljivke velja znani izrek.

Izrek 0.0.1 (Centralni limitni izrek). *Naj bodo $X_1, X_2, \ldots$ neodvisne in enako porazdeljene slučajne spremenljivke z matematičnimi upanji μ in variancami σ^2. Tedaj porazdelitve slučajnih spremenljivk:*

$$\frac{X_1 + \ldots X_n - n\mu}{\sqrt{n}}$$

šibko konvergirajo proti normalni porazdelitvi $\mathrm{N}(0, \sigma^2)$.

Veliko težje pa je vprašanje, koliko se porazdelitve naših vsot dejansko razlikujejo od normalne. V precejšnji meri nam odgovor na to da Berry–Esseenov izrek.

Izrek 0.0.2 (Berry, Esseen). *Naj bodo* $X_1, \ldots X_n$ *neodvisne in enako porazdeljene slučajne spremenljivke. Naj bo* $\mathbf{E}(X_k) = 0$ *in* $\operatorname{var}(X_k) = 1$. *Naj bo tudi* $\beta := \mathbf{E}(|X_k|^3) < \infty$. *Označimo* $X := n^{-1/2} \sum_{k=1}^n X_k$. *Tedaj velja:*

$$\sup_x \bigl|\mathbf{P}[X \le x] - \Phi(x)\bigr| \le \frac{C\beta}{\sqrt{n}}$$

Tu je C *pa je univerzalna konstanta, za katero je znano, da je* $\frac{1}{2\pi} \le C \le 0.8$.

Pri aproksimaciji s Poissonovo porazdelitvijo pa nam oceno napake da *Le Camov izrek* (glej [7]).

Izrek 0.0.3 (Le Cam). *Naj bodo* $X_1, \ldots X_n$ *neodvisne Bernoullijeve slučajne spremenljivke in* $\mathbf{P}[X_k = 1] = p_k$. *Naj bo* $X := X_1 + \ldots + X_n$ *in* $\lambda := p_1 + \ldots + p_n$. *Tedaj velja:*

$$\sup_{A \subset \mathbb{N}_0} \left|\mathbf{P}[X \in A] - \sum_{k \in A} \frac{\lambda^k e^{-\lambda}}{k!}\right| \le \min\left\{1, \tfrac{8}{\lambda}\right\} \sum_{k=1}^n p_k^2$$

Vsi trije izreki se na klasičen način dokažejo z uporabo karakterističnih funkcij. t. j. Fourierjevih transformacij porazdelitev. Vse temelji na dejstvu, da je karakteristična funkcija vsote *neodvisnih* slučajnih spremenljivk enaka produktu karakterističnih funkcij posameznih porazdelitev. Toda to orodje dobro deluje le pri neodvisnih slučajnih spremenljivkah.

Leta 1970 pa je profesor Charles Stein odkril popolnoma nov način ocenjevanja napake pri normalni aproksimaciji, ki je lepo deloval tudi pri rahlo odvisnih slučajnih spremenljivkah. Malo kasneje je profesor Louis H. Y. Chen odkril, da se da ta metoda uporabiti tudi pri Poissonovi aproksimaciji. Tako je nastala Chen–Steinova metoda. Lep pregled te metode nam da knjiga [24].

Pripomniti velja, da Chen–Steinova metoda nikakor ni omejena le na Poissonovo in normalno aproksimacijo. Tako npr. članek [17] obravnava aproksimacijo z multinomialno porazdelitvijo, delo [18] obravnava aproksimacijo s porazdelitvijo gama, delo [19] pa obravnava aproksimacijo z geometrijsko porazdelitvijo, vse po Chen–Steinovi metodi.

V tem delu se bomo omejili le na Poissonovo in normalno aproksimacijo, saj že ti dve naneseta več kot dovolj. V prvem poglavju obravnavamo osnovno idejo, to je, da porazdelitev μ, s katero aproksimiramo (npr. Poissonovo ali normalno) gledamo kot ravnovesno porazdelitev primernega markovskega (Feller–Dynkinovega) procesa. Če želimo porazdelitev dane slučajne spremenljivke X aproksimirati z μ, lahko ocenjujemo $\mathbf{E}(f(X)) - \int f\, d\mu$ za določen razred testnih funkcij. Pri Chen–Steinovi metodi pa ne ocenjujemo tega izraza, marveč izraz $\mathbf{E}((Gg)(X))$, kjer je G generator operatorske polgrupe markovskega procesa. Natančneje, velja:

$$\mathbf{E}(f(X)) - \int f\, d\mu = \mathbf{E}\bigl((Gg)(X)\bigr)$$

kjer je g rešitev *Steinove enačbe*:

$$Gg = f - \int f\,d\mu$$

Za primeren razred testnih funkcij f moramo torej rešiti Steinovi enačbo in nato čim več povedati o funkcijah g, da bomo znali dobro oceniti $\mathbf{E}((Gg)(X))$. Izkaže se, da moramo oceniti odvode ali diference funkcije g. Ta korak je odvisen le od porazdelitve, s katero aproksimiramo, in na ta korak je to delo v glavnem tudi osredotočeno. Omeniti velja, da ocenjevanje odvodov ali diferenc Steinove enačbe še zdaleč ni preprosto, vendar pa je, ko je enkrat narejeno, narejeno za vselej.

V prvem poglavju izpeljemo splošno pot, ki vodi do rešitve Steinove enačbe. Seveda pa ni nujno, da se te poti kasneje vselej držimo. V resnici se je držimo le v četrtem poglavju pri večrazsežni normalni aproksimaciji, medtem ko pri Poissonovi in enorazsežni normalni aproksimaciji Steinovo enačbo rešimo neposredno. Pripomniti pa velja, da se pomembna ocena pri Poissonovi aproksimaciji (namesto katere mi izpeljemo nekoliko slabšo, glej lemo 2.3.2 in drugo opombo k njej) izpelje prav po poti, opisani v prvem poglavju.

Ocenjevanje izraza $\mathbf{E}((Gg)(X))$ pa je seveda čisto odvisno od problema, ki ga rešujemo. V tem delu ta korak naredimo za neodvisne slučajne spremenljivke. Ocena za neodvisne slučajne spremenljivke pa je tudi osnova za splošnejše ocene, kjer so slučajne spremenljivke odvisne. V osnovni oceni je namreč vselej skrita slučajna spremenljivka X^*, katere porazdelitev je na določen način transformirana porazdelitev slučajne spremenljivke X. Če znamo slučajno spremenljivko X^* s tako transformirano porazdelitvijo najti tudi v primeru, ko je X vsota odvisnih slučajnih spremenljivk, in to na *istem verjetnostnem prostoru kot* X, nam to že da oceno za $\mathbf{E}((Gg)(X))$. Primer takega ocenjevanja je obdelan v drugem poglavju pri Poissonovi aproksimaciji (glej zgled na strani 19).

Delo ima dva dodatka. V prvem so opisane najpogostejše topologije in metrike na prostorih porazdelitev. Drugi dodatek pa predstavi markovske procese, ki tičijo v ozadju Chen–Steinove metode. Markovski procesi so predstavljeni precej podrobno. Morda je za to potrebno opravičilo, saj se da Chen–Steinova metoda obravnavati celo čisto brez njih. Za prej omenjeni operator G lahko vsaj pri Poissonovi in enorazsežni normalni aproksimaciji hitro pozabimo, da je generator Feller–Dynkinove polgrupe. Vendar pa potem nimamo več ideje, kako bi tak operator dobili še za kako drugo porazdelitev.

Zavedam se, da je v tem delu kljub pregledu verjetno ostala prenekatera napaka. Za vse morebitne napake se bralcu že vnaprej opravičujem. Prav tako se zavedam, da bi se dalo v to delo vključiti še precej več zanimivih primerov uporabe Chen–Steinove metode za odvisne slučajne spremenljivke. Več bi se dalo povedati tudi o večrazsežni normalni aproksimaciji, dalo pa bi se tudi na kratko obravnavati Chen–Steinovo metodo pri aproksimaciji z drugimi porazdelitvami. Vendar pa bi to preseglo obseg tega dela – predvsem rok, v katerem je moralo to delo nastati. Vseeno pa ponižno upam, da delo deluje kot zaključena celota in da se bodo bralci iz njega tudi kaj naučili.

V Ljubljani, 20. aprila 1998

Martin Raič

Najpogostejše oznake

$\int f(x)\,\mu(dx)$	Integral funkcije f po meri μ, t. j. $\int f\,d\mu$.
$f_*\mu$	Mera μ, preslikana s preslikavo f, t. j. $(f_*\mu)(A) := \mu(f^{-1}(A))$.
$\mathbf{P}$	Verjetnostna mera na temeljnem verjetnostnem prostoru.
$\mathcal{L}(X)$	Porazdelitev slučajne spremenljivke X, t. j. $X_*\mathbf{P}$.
$[\mathcal{A}]$	Dogodek, da se zgodi izjava $\mathcal{A}$, t. j. množica točk, na kateri je izjava $\mathcal{A}$ pravilna. Tako je npr. $[X \geq 0] = X^{-1}([0,\infty))$.
I_A	Indikatorska funkcija dogodka A, t. j. $I_A(x) = 1$, če je $x \in A$, drugače pa je $I_A(x) = 0$.
$I[\mathcal{A}]$	Indikator izjave $\mathcal{A}$, se pravi 1, če je izjava $\mathcal{A}$ pravilna, drugače pa 0.
$\mathrm{Pr}(S)$	Prostor vseh porazdelitev (= verjetnostnih mer) na S.
$\mathbb{N}$	$\{1, 2, 3, \ldots\}$.
$\mathbb{N}_0$	$\{0, 1, 2, \ldots\}$.
$X \sim \mu$	Slučajna spremenljivka X ima porazdelitev μ, t. j. $\mu = \mathcal{L}(X)$.
$\mathrm{Bin}(n, p)$	Binomska porazdelitev.
$\mathrm{Po}(\lambda)$	Poissonova porazdelitev.
$\mathrm{N}(a, \sigma^2)$	Normalna (Gaussova) porazdelitev. Pripomniti velja, da drugi parameter predstavlja varianco in ne standardnega odklona, kar je usklajeno z oznako za večrazsežno normalno porazdelitev (glej spodaj).
$\mathrm{N}(a, \Sigma)$	Večrazsežna normalna porazdelitev. Tu je Σ kovariančna matrika.
γ_p	Standardna normalna porazdelitev v $\mathbb{R}^p$, t. j. $\mathrm{N}(0, I)$.
$\Phi(x)$	Verjetnostni integral: $\Phi(x) := \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt$
$\widehat{S}$	Kompaktifikacija topološkega prostora S z eno točko.
$\mathcal{C}(S)$	Prostor zveznih realnih funkcij na topološkem prostoru S.
$\mathcal{C}_0(S)$	Prostor zveznih realnih funkcij na S, ki gredo v neskončnosti proti 0, t. j. zunaj dovolj velike kompaktne množice so po absolutni vrednosti poljubno majhne.
$\mathcal{C}_\kappa(S)$	Prostor zveznih realnih funkcij na S s kompaktnim nosilcem.
$\mathcal{C}^{(k)}(\mathbb{R}^p)$	Prostor k-krat zvezno odvedljivih funkcij na $\mathbb{R}^p$.
$\mathcal{C}_L(D)$	Prostor zveznih preslikav iz D v L.
$\mathcal{C}_\flat(S)$	Prostor omejenih merljivih realnih funkcij na merljivem prostoru S (t. j. s podano σ-algebro, brez mere).

1.

Izpeljava Chen–Steinove metode

Chen–Steinova metoda nam pomaga oceniti razdaljo med porazdelitvijo ν, ki jo dobimo kot porazdelitev kake zapletene slučajne spremenljivke, in 'limitno' porazdelitvijo μ. Porazdelitev μ je navadno Poissonova ali normalna, vendar pa Chen–Steinova metoda nikakor ni omejena le na aproksimacijo s tema dvema porazdelitvama.

Osnovna ideja Chen–Steinove metode je, da konstruiramo Feller-Dynkinov proces z ravnovesno porazdelitvijo μ. Pri Poissonovi porazdelitvi je to primeren proces rojevanja in umiranja, pri normalni pa je to Ornstein–Uhlenbeckov proces. Pripomnimo naj, da za dano porazdelitev lahko obstaja več primernih slučajnih procesov, ki imajo to porazdelitev za ravnovesno.

Naj bo torej μ ravnovesna porazdelitev Feller–Dynkinovega procesa na lokalno kompaktnem Hausdorffovem prostoru S s števno bazo topologije. Procesu naj pripada operatorska polgrupa $(P_t)_{t\in[0,\infty)}$ z generatorjem G (glej dodatek B). Če za začetno porazdelitev postavimo neko porazdelitev ν in opazujemo, kako se spreminja s časom, se pri dobrem procesu začnejo robne porazdelitve vselej približevati invariantni porazdelitvi μ. Pri Chen–Steinovi metodi gledamo, *kako hitro* se začne porazdelitev ν bližati porazdelitvi μ in na podlagi tega ocenjujemo razdaljo med μ in ν.

Hitrost približevanja pa ni nič drugega kot delovanje generatorja. Natančneje, če je X markovski proces z začetno porazdelitvijo ν in jedrom, ki ima G za generator, je za primerno funkcijo g integral $\int Gg\,d\nu$ ravno odvod funkcije $t \mapsto \mathbf{E}(g(X_t))$ v točki 0. Če je $\nu = \mu$, je $\mathbf{E}(g(X_t))$ seveda konstanta. Pri dobrem procesu bo $\nu = \mu$ natanko tedaj, ko bo $\int Gg\,d\nu = 0$ za vsako funkcijo $g \in \mathcal{D}(G)$ (glej trditev B.14.2). Če bomo torej za primerno družino testnih funkcij g dobro ocenili izraz $\int Gg\,d\nu$, bomo morda dobro ocenili razdaljo med μ in ν.

Razdaljo med μ in ν bomo ocenili tako, da bomo za primeren razred testnih funkcij f ocenili $\int f\,d\nu - \int f\,d\mu$. Če za f vzamemo vse merljive funkcije, ki so po absolutni vrednosti navzgor omejene z 1, dobimo oceno v metriki totalne variacije, če pa vzamemo vse neraztezne funkcije (t. j. Lipschitzove funkcije s konstanto 1), dobimo oceno v Wassersteinovi metriki (glej dodatek A, stran 76). Razliko integralov pa bomo izrazili kot integral $\int Gg\,d\nu$. Velja namreč:

$$\int f\,d\nu - \int f\,d\mu = \int Gg\,d\nu$$

kjer je g rešitev *Steinove enačbe*:

$$Gg = f - \int f \, d\mu \tag{1.0.1}$$

Če je generator G dovolj preprost, lahko Steinovo enačbo rešimo povsem elementarno. V splošnem pa se njena rešitev izraža s pomočjo operatorske polgrupe $(P_t)_{t\in[0,\infty)}$, ki jo generira G, toda le v primeru, ko proces dovolj hitro konvergira k ravnovesni porazdelitvi μ. Natančneje, za vsako testno funkcijo f mora iti $P_t f$ dovolj hitro proti $\int f \, d\mu$, ko gre t proti neskončno. Funkcija $P_t f$ je namreč regresija slučajne spremenljivke $f(X_t)$ glede na X_0 in pri dobrem procesu in dovolj velikem t ima X_t porazdelitev blizu μ in je skoraj neodvisna od X_0.

Videli bomo, da bodo morale iti funkcije $P_t f$ iti proti $\int f \, d\mu$ enakomerno, in to dovolj hitro. To pa navadno ne bo res kar v prostoru $\mathcal{C}_0(S)$, v katerem navadno gledamo operatorje P_t. Nerodno je že to, da v primeru, ko S ni kompakten, konstanta $\int f \, d\mu$ ne pripada $\mathcal{C}_0(S)$. Zato operatorje P_t razširimo na kak večji prostor L, ki bo moral med drugim vsebovati tudi konstante. Prostor L bo treba opremiti s primerno normo, v kateri bo poln. Pripomniti pa velja, da norma na L na preseku $L \cap \mathcal{C}_0(S)$ ne bo nujno inducirala supremum norme.

Operatorje P_t lahko gledamo tudi na prostoru $\flat(S, \mathcal{S})$ omejenih merljivih funkcij iz S v $\mathbb{R}$, kjer je $\mathcal{S}$ Borelova σ-algebra. Novo operatorsko polgrupo na L bomo definirali tako, da se bodo stari operatorji na preseku $\flat(S, \mathcal{S}) \cap L$ ujemali z novimi.

Operatorska polgrupa na $\mathcal{C}_0(S)$ je krepko zvezna in kontrakcijska (glej razdelka o Feller–Dynkinovih procesih in generatorjih operatorskih polgrup v dodatku B). Za novo operatorsko polgrupo na L bo seveda treba posebej dokazati, da so njeni elementi omejeni operatorji (kontrakcije v splošnem ne bodo) in da je polgrupa krepko zvezna. Poleg tega pa bo seveda treba še dokazati, da gredo funkcije $P_t f$ dovolj hitro proti $\int f \, d\mu$.

Ključni korak k rešitvi Steinove enačbe je trditev B.11.7. Iz nje sledi, da za vsako funkcijo $f \in \mathcal{D}(G)$ velja *Dynkinova formula*:

$$P_t f - f = \int_0^t G P_s f \, ds$$

Pri dobrem procesu z ravnovesno porazdelitvijo μ in dobro izbranem prostoru L bo izraz $P_t f$ po normi konvergiral h konstanti $c := \int f \, d\mu$, ko bo šel čas t proti neskončno. Levi izraz bo torej konvergiral k $c - f$. Potem pa bo konvergiral tudi desni izraz in po definiciji posplošenega Riemannovega integrala (glej stran 115) bo veljalo:

$$c - f = \int_0^\infty G P_s f \, ds$$

Če bo šel izraz $P_t f$ dovolj hitro proti c, pa bo obstajal tudi naslednji posplošeni Riemannov integral:

$$\int_0^\infty \big(P_s f - c\big) \, ds$$

Ta integral moramo zdaj preslikati z generatorjem G. Ker je $f \in \mathcal{D}(G)$, je po trditvi B.11.3 tudi $P_s f \in \mathcal{D}(G)$. Ker operatorji P_t ohranjajo konstante, morajo biti tudi le-te v definicijskem območju generatorja G in ta jih mora preslikati v 0. Po posledici B.11.8 je generator G zaprt. Po točki (2) trditve B.11.6 potem tudi zgornji posplošeni Riemannov integral pripada $\mathcal{D}(G)$ in velja:

$$G \int_0^\infty \big(P_s f - c\big)\, ds = \int_0^\infty G\big(P_s f - c\big)\, ds = \int_0^\infty G P_s f\, ds = c - f$$

Dokazali smo naslednji izrek.

Izrek 1.0.1. *Naj bo $(P_t)_{t\in[0,\infty)}$ krepko zvezna operatorska polgrupa na Banachovem prostoru L z generatorjem G. Naj bo $f \in L$ in naj gre $P_t f$ proti elementu $c \in L$, in sicer tako hitro, da velja:*

$$\int_0^\infty \big\|P_s f - c\big\|\, ds < \infty$$

Za vsak $t \in [0, \infty)$ naj bo tudi $P_t c = c$. Tedaj rešitev Steinove enačbe $Gg = f - c$ obstaja in se izraža v obliki:

$$g = -\int_0^\infty \big(P_s f - c\big)\, ds$$

Enačba (1.0.1) ima torej rešitev oblike:

$$g = -\int_0^\infty \left(P_s f - \int f\, d\mu\right) ds$$

Če enačbo (1.0.1) integriramo po porazdelitvi ν, dobimo, da za tako definirano funkcijo g velja:

$$\int f\, d\nu - \int f\, d\mu = \int Gg\, d\nu$$

V dosti primerih se bo izkazalo, da je $\int Gg\, d\nu$ laže oceniti kot $\int f\, d\nu - \int f\, d\mu$.

Osnovni recept je torej takle: vzamemo primerno družino testnih funkcij f in skušamo čim več povedati o funkcijah g, ki so pripadajoče rešitve Steinove enačbe. Ta korak je odvisen le od porazdelitve μ in procesa, ki smo ga izbrali. Drugi korak pa je ocena integrala $\int Gg\, d\nu$ iz parametrov, ki smo jih ocenili v prvem koraku. Ta pa je seveda odvisna tudi od porazdelitve ν, torej od problema, ki ga rešujemo.

Kako sploh dobiti primeren markovski proces, ki ima μ za invariantno porazdelitev? Eden od načinov je, da izhajamo kar iz problema, ki ga rešujemo, in naredimo najprej *drugi korak*. Slučajni spremenljivki Y, katere porazdelitev aproksimiramo, poiščemo podobno slučajno spremenljivko Y' na istem verjetnostnem prostoru in z isto porazdelitvijo, pri čemer verjetnostni prostor, na katerem je definirana Y, po potrebi razširimo. Nato za primerno družino testnih funkcij g razvijemo izraz $\mathbf{E}(g(Y'))$ in če imamo srečo, dobimo rezultat oblike:

$$\mathbf{E}(g(Y')) = \mathbf{E}(g(Y)) + t\mathbf{E}((Gg)(Y)) + t\varepsilon(g)$$

kjer je G generator Feller–Dynkinovega procesa. Tako dobimo oceno:

$$|\mathbf{E}((Gg)(Y))| = \left|\int Gg\, d\mathcal{L}(Y)\right| \leq |\varepsilon(g)|$$

Drugi korak je tako narejen. Zdaj, ko imamo generator, pa je treba le še poiskati rešitev Steinove enačbe in povedati čim več ocen o njej.

2.

Poissonova aproksimacija

2.1 Motivacija

Naj bodo $X_1, \ldots X_n$ neodvisne Bernoullijeve slučajne spremenljivke, t. j. $\mathbf{P}[X_k = 0] = 1 - p_k$ in $\mathbf{P}[X_k = 1] = p_k$. Želeli bi aproksimirati porazdelitev vsote $X := X_1 + \ldots + X_n$. V skladu s prejšnjim poglavjem bomo to storili tako, da bomo verjetnostni prostor, na katerem so definirane naše slučajne spremenljivke, še malo razširili in poiskali slučajno spremenljivko X', ki bo malo drugačna kot X, porazdeljena pa bo enako. Njeno konstrukcijo povzemamo iz [3].

Izberimo neodvisne slučajne spremenljivke K in $Y_1, \ldots Y_n$ takole: slučajna spremenljivka K naj bo porazdeljena enakomerno na $\{1, \ldots n\}$ in za vsak k naj bo Y_k porazdeljena tako kot X_k. Slučajne spremenljivke K, $Y_1, \ldots Y_n$ naj bodo neodvisne in tudi neodvisne od $X_1, \ldots X_n$. Definirajmo $D_k := Y_k - X_k$ in naj bo $X' := X + D_K$. Tako slučajna spremenljivka X' ni preveč različna od X, če so le parametri p_k dovolj majhni. Trdimo: X' je enako porazdeljena kot X. Velja namreč:

$$
\begin{aligned}
\mathbf{P}[X' \in A, K = k] &= \mathbf{P}[X + D_k \in A, K = k] = \\
&= \mathbf{P}[X_1 + \ldots + X_{k-1} + Y_k + X_{k+1} + \ldots + X_n \in A, K = k] = \\
&= \mathbf{P}[X_1 + \ldots + X_{k-1} + Y_k + X_{k+1} + \ldots + X_n \in A]\, \mathbf{P}[K = k] = \\
&= \mathbf{P}[X_1 + \ldots + X_n \in A]\, \mathbf{P}[K = k]
\end{aligned}
\tag{2.1.1}
$$

Upoštevali smo neodvisnost in dejstvo, da sta X in $X + D_k$ enako porazdeljeni, saj sta obe slučajni spremenljivki vsoti neodvisnih Bernoullijevih slučajnih spremenljivk s parametri p_i. Ko zdaj vse skupaj seštejemo po k, dobimo, da je X' res porazdeljena enako kot X.

Naj bo zdaj g poljubna funkcija iz $\mathbb{N}_0$ v $\mathbb{R}$. Tedaj velja:

$$
\begin{aligned}
\mathbf{E}\big(g(X)\big) = \mathbf{E}\big(g(X')\big) &= \sum_{k=1}^{n} \Big(\mathbf{E}\big(g(X')\, I_{[Y_k=0]}\, I_{[K=k]}\big) + \mathbf{E}\big(g(X')\, I_{[Y_k=1]}\, I_{[K=k]}\big) \Big) = \\
&= \sum_{k=1}^{n} \Big(\mathbf{E}\big(g(X - X_k)\, I_{[Y_k=0]}\, I_{[K=k]}\big) + \mathbf{E}\big(g(X - X_k + 1)\, I_{[Y_k=1]}\, I_{[K=k]}\big) \Big) = \\
&= \frac{1}{n} \sum_{k=1}^{n} \Big((1 - p_k) \mathbf{E}\big(g(X - X_k)\big) + p_k \mathbf{E}\big(g(X - X_k + 1)\big) \Big)
\end{aligned}
$$

Zgolj iz formalnih razlogov izberimo vrednost $g(-1)$, ki na koncu ne bo igrala nobene vloge, in za $n \in \mathbb{N}_0$ definirajmo:

$$(\Delta g)(n) := g(n) - g(n-1)$$

Ker X_k lahko zavzame le vrednosti 0 in 1, velja:

$$\mathbf{E}\big(g(X)\big) = \frac{1}{n}\sum_{k=1}^{n}\Big((1-p_k)\big(g(X) - X_k(\Delta g)(X)\big) + p_k\big(g(X+1) - X_k(\Delta g)(X+1)\big)\Big)$$

Od obeh strani odštejmo $E\big(g(X)\big)$ in pomnožimo z n. Po krajšem računu dobimo:

$$\begin{aligned}0 &= \mathbf{E}\Big(p_k(\Delta g)(X+1) - X_k(\Delta g)(X) - p_k X_k(\Delta^2 g)(X+1)\Big) = \\ &= \mathbf{E}\big(\lambda(\Delta g)(X+1) - X(\Delta g)(X)\big) - \sum_{k=1}^{n} p_k \mathbf{E}\big(X_k(\Delta^2 g)(X+1)\big)\end{aligned}$$

Označili smo $\lambda := p_1 + \ldots p_n$. Če označimo:

$$M_2(g) := \sup_{r\geq 2}\big|(\Delta^2 g)(r)\big|$$

velja ocena:

$$\Big|\mathbf{E}\big(X_k(\Delta^2 g)(X+1)\big)\Big| \leq \mathbf{E}(X_k)\, M_2(g) = p_k\, M_2(g)$$

(če je $X = 0$, je tudi $X_k = 0$ in ocena velja ne glede na to, koliko je $g(-1)$). Sledi:

$$\Big|\mathbf{E}\big(\lambda(\Delta g)(X+1) - X(\Delta g)(X)\big)\Big| \leq M_2(g)\sum_{k=1}^{n} p_k^2$$

Izraz pod absolutno vrednostjo je enak $\mathbf{E}\big((Gg)(X)\big)$, kjer je G generator procesa rojevanja in umiranja s koeficienti $a_n = \lambda$ in $b_n = n$ (glej stran 120). Ta proces ima invariantno porazdelitev $\mathrm{Po}(\lambda)$ (glej stran 137). Rešiti moramo torej Steinovo enačbo:

$$Gg = f - \int f\, d\,\mathrm{Po}(\lambda) \tag{2.1.2}$$

pa bomo imeli oceno:

$$\left|\mathbf{E}\big(f(X)\big) - \int f\, d\,\mathrm{Po}(\lambda)\right| \leq M_2(g)\,\sum_{k=1}^{n} p_k^2$$

Izraz $(Gg)(r) = \lambda(\Delta g)(r+1) - r(\Delta g)(r)$ se izraža le z Δg, prav tako tudi zgornja ocena. Zato lahko zgornjo ugotovitev zapišemo tudi takole: če je h rešitev enačbe:

$$\lambda h(r+1) - r h(r) = f(r) - \int f\, d\,\mathrm{Po}(\lambda)$$

velja ocena:

$$\left|\mathbf{E}\big(f(X)\big) - \int f\, d\,\mathrm{Po}(\lambda)\right| \leq M_1(h)\,\sum_{k=1}^{n} p_k^2$$

kjer je $M_1(h) = \sup_{r\geq 2}\big|(\Delta h)(r)\big|$. Več o tem pa v naslednjem razdelku.

2.2 Ocenjevanje z drugimi diferencami

V prejšnjem razdelku smo dokazali naslednjo lemo.

Lema 2.2.1. *Naj bodo $X_1, \dots X_n$ neodvisne Bernoullijeve slučajne spremenljivke in $\mathbf{P}[X_k = 1] = p_k$. Naj bo $X := \sum_{k=1}^n X_k$ in $\lambda := \sum_{k=1}^n p_k$. Tedaj za vsako funkcijo $h\colon \mathbb{N}_0 \to \mathbb{R}$ velja ocena:*

$$\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big) \le M_1(h) \sum_{k=1}^n p_k^2$$

Prejšnji razdelek pa je služil bolj za to, da smo dobili občutek, zakaj gledati ravno $\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big)$. Lema je bila zato izpeljana na bistveno daljši način, kot je v resnici potrebno. Zato bomo navedli še krajši dokaz, ki bo tudi osnova vseh nadaljnjih ocen pri Poissonovi aproksimaciji.

Krajši dokaz leme. Označimo $W_k := X - X_k$. Velja:

$$\begin{aligned}
&\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big) = \\
&\quad = \sum_{k=1}^n \mathbf{E}\Big(p_k h(X+1) - X_k h(X)\Big) = \\
&\quad = \sum_{k=1}^n \mathbf{E}\Big(p_k h(X+1) - X_k h(W_k+1)\Big) = \\
&\quad = \sum_{k=1}^n \mathbf{E}\Big(p_k(1-X_k)h(W_k+1) + p_k X_k h(W_k+2) - X_k h(W_k+1)\Big) = \\
&\quad = \sum_{k=1}^n \mathbf{E}\Big(p_k(1-p_k)h(W_k+1) + p_k^2 h(W_k+2) - p_k h(W_k+1)\Big) = \\
&\quad = \sum_{k=1}^n p_k^2 \mathbf{E}\big((\Delta h)(W_k+2)\big) \le M_1(h) \sum_{k=1}^n p_k^2
\end{aligned}$$

■

Rešiti je treba še prirejeno Steinovo enačbo:

$$\lambda h(r+1) - rh(r) = f(r) - \int f \, d\,\mathrm{Po}(\lambda) \tag{2.2.1}$$

in oceniti $M_1(h)$. Oceniti je torej treba druge diference rešitve g originalne Steinove enačbe (2.1.2), ki so natančno prve diference funkcije h.

Opomba. Pri rekurzivnih enačbah, kakršna je tudi (2.2.1), moramo za enoličnost rešitve postaviti še začetni pogoj, ki je v našem primeru vrednost $h(0)$. Vendar pa v našem primeru nadaljnje vrednosti funkcije h (in z njimi tudi vrednost $M_1(h)$) niso prav nič odvisne od $h(0)$. Začetni pogoj torej ne igra nobene vloge.

Idejo ocene povzemamo iz [6] (glej stran 8). Postavimo:

$$\varphi := f - \int f \, d\,\mathrm{Po}(\lambda) \tag{2.2.2}$$

za $r = 1, 2, \ldots$ pa postavimo:

$$\psi(r) := \frac{\lambda^r}{(r-1)!} h(r)$$

in dodajmo še $\psi(0) := 0$. Tedaj za $r = 0, 1, \ldots$ velja:

$$\psi(r+1) - \psi(r) = \frac{\lambda^r}{r!} \varphi(r)$$

Sledi:

$$\psi(r) = \sum_{i=0}^{r-1} \frac{\lambda^i}{i!} \varphi(i)$$

Za $r = 1, 2, \ldots$ torej velja:

$$h(r) = \frac{(r-1)!}{\lambda^r} \sum_{i=0}^{r-1} \frac{\lambda^i}{i!} \varphi(i) \tag{2.2.3}$$

Zdaj, ko smo rešili enačbo (2.2.1), pa moramo oceniti še $M_1(h)$. Najprej za vsak $r \in \mathbb{N}$ velja:

$$\begin{aligned} h(r+1) - h(r) &= \frac{r!}{\lambda^{r+1}} \sum_{i=0}^{r} \frac{\lambda^i}{i!} \varphi(i) - \frac{(r-1)!}{\lambda^r} \sum_{i=0}^{r-1} \frac{\lambda^i}{i!} \varphi(i) = \\ &= \frac{r!}{\lambda^{r+1}} \varphi(0) + \frac{r!}{\lambda^{r+1}} \sum_{i=1}^{r} \frac{\lambda^i}{i!} \varphi(i) - \frac{(r-1)!}{\lambda^{r+1}} \sum_{i=1}^{r} \frac{\lambda^i}{(i-1)!} \varphi(i-1) = \\ &= \frac{r!}{\lambda^{r+1}} \varphi(0) + \frac{(r-1)!}{\lambda^{r+1}} \sum_{i=1}^{r} \frac{\lambda^i}{i!} \big(r\varphi(i) - i\varphi(i-1) \big) \end{aligned} \tag{2.2.4}$$

Formula (2.2.3) nam pove, da je vrednost $h(r)$ odvisna le od $\varphi(0), \ldots \varphi(r-1)$. Potrebovali pa bomo tudi, da je vrednost $h(r)$ odvisna tudi le od $\varphi(r), \varphi(r+1), \ldots$. To sledi iz enakosti $\int \varphi \, d\,\mathrm{Po}(\lambda) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \varphi(i) = 0$. Dobimo:

$$h(r) = -\frac{(r-1)!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{i!} \varphi(i) \tag{2.2.5}$$

Sledi:

$$\begin{aligned} h(r+1) - h(r) &= \frac{(r-1)!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{i!} \varphi(i) - \frac{r!}{\lambda^{r+1}} \sum_{i=r+1}^{\infty} \frac{\lambda^i}{i!} \varphi(i) = \\ &= \frac{(r-1)!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{i!} \varphi(i) - \frac{r!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{(i+1)!} \varphi(i) = \\ &= \frac{(r-1)!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{(i+1)!} \big((i+1)\varphi(i) - r\varphi(i+1) \big) \end{aligned} \tag{2.2.6}$$

Kako je porazdelitev slučajne spremenljivke X blizu Poissonovi porazdelitvi $\mathrm{Po}(\lambda)$, bomo ocenjevali v metriki totalne variacije, torej bomo za testne funkcije f postavljali indikatorje. Za vsako množico $A \subset \mathbb{N}_0$ naj bo h_A rešitev enačbe (2.2.1) za $f = I_A$. Oceniti moramo $M_1(h_A)$. Najprej bomo ocenili $M_1(h_{\{j\}})$ za $j = 0, 1, \ldots$. Za vse $r = 1, 2, \ldots$ je torej treba oceniti $h_{\{j\}}(r+1) - h_{\{j\}}(r)$. Naj bo najprej $r < j$. Tedaj je $\varphi(i) = -\frac{\lambda^j}{j!}e^{-\lambda}$ za vse $i = 0, \ldots r$. Iz verige neenakosti (2.2.4) sledi:

$$h_{\{j\}}(r+1) - h_{\{j\}}(r) = -\frac{\lambda^j}{j!}e^{-\lambda}\left(\frac{r!}{\lambda^{r+1}} + \frac{(r-1)!}{\lambda^{r+1}}\sum_{i=1}^{r}\frac{\lambda^i}{i!}(r-i)\right) \leq 0$$

V primeru, ko je $r = j$, prav tako iz (2.2.4) dobimo:

$$h_{\{j\}}(j+1) - h_{\{j\}}(j) = -\frac{\lambda^j}{j!}e^{-\lambda}\left(\frac{j!}{\lambda^{j+1}} + \frac{(j-1)!}{\lambda^{j+1}}\sum_{i=1}^{j}\frac{\lambda^i}{i!}(j-i)\right) + \frac{(j-1)!}{\lambda^{j+1}}\frac{\lambda^j}{j!}j$$

Drugi člen izraza v oklepaju je večji ali enak 0. Če ga spustimo, dobimo oceno:

$$h_{\{j\}}(j+1) - h_{\{j\}}(j) \leq \frac{1-e^{-\lambda}}{\lambda}$$

Če je $r > j$, pa enakost $\varphi(i) = -\frac{\lambda^j}{j!}e^{-\lambda}$ velja za $i = r, r+1, \ldots$. Sledi:

$$h_{\{j\}}(r+1) - h_{\{j\}}(r) = -\frac{\lambda^j}{j!}e^{-\lambda}\frac{(r-1)!}{\lambda^r}\sum_{i=r}^{\infty}\frac{\lambda^i}{(i+1)!}(i+1-r) \leq 0$$

Dobili smo, da za vsak $r \in \mathbb{N}$ in vsak $j \in \mathbb{N}_0$ velja:

$$\begin{aligned} h_{\{j\}}(r+1) - h_{\{j\}}(r) &\leq 0, \quad \text{če je } j \neq r \\ h_{\{j\}}(j+1) - h_{\{j\}}(j) &\leq \frac{1-e^{-\lambda}}{\lambda} \end{aligned} \tag{2.2.7}$$

Ocenili smo $h_{\{j\}}$, zdaj pa ocenimo še h_A. Trdimo:

$$h_A = \sum_{j\in A} h_{\{j\}} \tag{2.2.8}$$

Dokazali bomo konvergenco po točkah. Najprej s φ_A označimo funkcijo φ v enačbi (2.2.2), kjer postavimo $f := I_A$. Z drugimi besedami, velja $\varphi_A(r) = I_A(r) - \mathrm{Po}(\lambda)(A)$. Tedaj očitno za vsako množico A velja $\varphi_A = \sum_{j\in A}\varphi_{\{j\}}$. V primeru, ko je A neskončna, vrsta konvergira po točkah. Iz formule (2.2.3) pa je jasno, da je vrednost $h_A(r)$ odvisna le od $\varphi_A(0), \ldots \varphi_A(r-1)$, torej vrsta po točkah konvergira tudi v formuli (2.2.8).

Iz formule (2.2.8) zdaj dobimo, da je za oceno diferenc funkcije h_A treba le sešteti ocene (2.2.7) po vseh $j \in A$. Pri tem je največ en člen pozitiven, pa še ta je navzgor omejen z $\frac{1-e^{-\lambda}}{\lambda}$. Sledi:

$$h_A(r+1) - h_A(r) \leq \frac{1-e^{-\lambda}}{\lambda}$$

Če postavimo $f = 1$, je $\varphi = 0$ in tudi $h = 0$. Zato je $h_{A^c} = -h_A$. Dobimo:

$$h_A(r+1) - h_A(r) \geq -\frac{1-e^{-\lambda}}{\lambda}$$

Dokazali smo naslednjo oceno drugih diferenc rešitve Steinove enačbe.

Lema 2.2.2. *Naj bo $\lambda > 0$, $A \subset \mathbb{N}_0$ in naj bo $h\colon \mathbb{N}_0 \to \mathbb{R}$ rešitev prirejene Steinove enačbe:*

$$\lambda h(r+1) - h(r) = I_A(r) - \mathrm{Po}(\lambda)(A) \tag{2.2.9}$$

Tedaj velja ocena:

$$M_1(h) \leq \frac{1-e^{-\lambda}}{\lambda}$$

■

Opomba. *Za $\lambda > 0$ velja* $\frac{1}{\lambda}\big(1-e^{-\lambda}\big) \leq \min\big\{1, \frac{1}{\lambda}\big\}$. Da je izraz manjši ali enak $\frac{1}{\lambda}$, je očitno. Trditev, da je izraz enak največ 1, pa je ekvivalentna neenačbi $1 - e^{-\lambda} \leq \lambda$. Le-ta pa drži, saj je $1 - e^{-\lambda} = \int_0^\lambda e^{-t}\,dt \leq \int_0^\lambda dt = \lambda$.

Če v lemi 2.2.2 v funkciji f in h vstavimo slučajno spremenljivko in izračunamo matematično upanje, dobimo naslednji izrek.

Izrek 2.2.3. *Naj bo X slučajna spremenljivka z vrednostmi v končni podmnožici množice $\mathbb{N}_0$, naj bo $\lambda > 0$ in naj za vsako funkcijo $h\colon \mathbb{N}_0 \to \mathbb{R}$ velja:*

$$\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big) \leq C M_1(h)$$

Tedaj velja:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \leq C\frac{1-e^{-\lambda}}{\lambda} \leq C \min\big\{1, \tfrac{1}{\lambda}\big\}$$

kjer je d_{TV} metrika totalne variacije, definirana na strani 72.

Pri neodvisnih slučajnih spremenljivkah nam je pri preverjanju pogojev zgornjega izreka na voljo lema 2.2.1. Dobimo naslednji izrek.

Izrek 2.2.4 (Le Cam). *Naj bodo $X_1, \ldots X_n$ neodvisne Bernoullijeve slučajne spremenljivke in $\mathbf{P}[X_k = 1] = p_k$. Naj bo $X := X_1 + \ldots + X_n$ in $\lambda := p_1 + \ldots + p_n$. Tedaj velja:*

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \leq \frac{1-e^{-\lambda}}{\lambda}\sum_{k=1}^n p_k^2 \leq \min\big\{1, \tfrac{1}{\lambda}\big\}\sum_{k=1}^n p_k^2$$

Če smo natančni, je ta izrek izboljšava Le Camovega izreka, pri čemer pa je izboljšana le konstanta. Le Cam namreč v [7] dokaže oceno:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \leq \min\big\{1, \tfrac{8}{\lambda}\big\}\sum_{k=1}^n p_k^2$$

Pripomnimo pa naj, da je Le Cam svoj izrek dokazal s popolnoma drugačnimi sredstvi.

Zgled 1. V primeru, ko so vsi parametri p_k enaki, dobimo naslednjo oceno:

$$d_{\mathrm{TV}}\big(\mathrm{Bin}(n,p),\mathrm{Po}(np)\big) \le p\left(1-e^{-np}\right) \le p\min\{1,np\} \tag{2.2.10}$$

Drugače povedano, za vsak $\lambda > 0$ zaporedje binomskih porazdelitev $\mathrm{Bin}\big(n,\frac{\lambda}{n}\big)$ konvergira proti Poissonovi porazdelitvi $\mathrm{Po}(\lambda)$ in hitrost konvergence je velikostnega reda $\frac{1}{n}$ ali hitrejša. No, v resnici je velikostni red $\frac{1}{n}$ kar pravi, saj velja:

$$\mathrm{Bin}\big(n,\tfrac{\lambda}{n}\big)(\{0\}) - \mathrm{Po}(\lambda)(\{0\}) = \left(1-\frac{\lambda}{n}\right)^n - e^{-\lambda}$$

Če definiramo funkcijo f po predpisu:

$$f(x) := \big(1-\lambda x\big)^{\frac{1}{x}}$$

in dodamo $f(0) := e^{-\lambda}$, dobimo zvezno funkcijo na realni osi in po Lagrangeovem izreku velja:

$$\mathrm{Bin}\big(n,\tfrac{\lambda}{n}\big)(\{0\}) - \mathrm{Po}(\lambda)(\{0\}) = f\big(\tfrac{1}{n}\big) - f(0) = \tfrac{1}{n}f'(\xi)$$

Ker je:

$$\begin{aligned} f'(x) &= -\big(1-\lambda x\big)^{\frac{1}{x}}\left(\frac{\log(1-\lambda x)}{x^2} + \frac{\lambda}{x(1-\lambda x)}\right) = \\ &= -\big(1-\lambda x\big)^{\frac{1}{x}}\Big(\tfrac{1}{2}\lambda^2 + \tfrac{2}{3}\lambda^3 x + \tfrac{3}{4}\lambda^4 x^2 + \dots\Big) \end{aligned}$$

za velike n velja:

$$\Big|\mathrm{Bin}\big(n,\tfrac{\lambda}{n}\big)(\{0\}) - \mathrm{Po}(\lambda)(\{0\})\Big| \approx \frac{\lambda^2 e^{-\lambda}}{2n}$$

Hitrost konvergence je torej res velikostnega reda $\frac{1}{n}$.

Opomba. Zgornja meja v oceni v lemi 2.2.2 je dosežena, če postavimo $A := \{1\}$. Velja namreč $h_{\{1\}}(2) - h_{\{1\}}(1) = \frac{1-e^{-\lambda}}{\lambda}$. Vendar pa to še zdaleč ne pomeni, da je ocena v prejšnji posledici najboljša možna, saj ocena v lemi 2.2.1 *ni* dosežena. Z bolj sofisticiranimi tehnikami, ki temeljijo na oceni supremum norme funkcije h, ki je rešitev enačbe (2.2.9) (torej ocene *prvih*, ne drugih diferenc rešitve Steinove enačbe), je npr. možno dokazati naslednjo oceno, ki je izpeljana v [25]:

$$d_{\mathrm{TV}}\big(\mathrm{Bin}(n,p),\mathrm{Po}(np)\big) \le \frac{0.6844\,p\sqrt{np}}{\sqrt{\lfloor np\rfloor\left(1-\frac{\lfloor np\rfloor}{n-1}\right)}} \tag{2.2.11}$$

Tu $\lfloor\cdot\rfloor$ pomeni celi del. Če je np velik, je ta ocena boljša od ocene (2.2.10).

Chen–Steinova metoda nam je dala lepo oceno za razdaljo med binomsko porazdelitvijo, ki je porazdelitev vsote neodvisnih slučajnih spremenljivk, in Poissonovo porazdelitvijo. Ena glavnih prednosti te metode pa je, da ni omejena le na neodvisne slučajne spremenljivke. V naslednjem zgledu bomo pokazali, da se da izrek 2.2.3 koristno uporabiti tudi v primeru, ko so slučajne spremenljivke rahlo odvisne.

Zgled 2. Ta zgled povzemamo iz [3]. Dana naj bo matrika $a(i,j)$, $1 \le i,j \le n$, z elementi, ki so ničle in enice. Matriko si lahko predstavljamo tudi kot usmerjen graf na n točkah. Naj bo π permutacija n elementov. Definirajmo:

$$X(\pi) := \sum_{i=1}^{n} a(i,\pi(i))$$

Če prostor Ω vseh permutacij n elementov opremimo z verjetnostno mero $\mathbf{P}$, ki naj bo kar enakomerna porazdelitev, postane X slučajna spremenljivka na Ω. Lahko si preprosto predstavljamo, da permutacija π postane slučajna. Tedaj je X še vedno oblike $\sum_{i=1}^{n} X_i$, kjer postavimo $X_i(\pi) := a(i,\pi(i))$. Slučajne spremenljivke X_i so sicer Bernoullijeve, niso pa več neodvisne. Toda če je število enic v matriki razmeroma majhno in so primerno razpršene, se da porazdelitev slučajne spremenljivke X še vedno dobro aproksimirati s Poissonovo. Natančneje, velja:

$$\mathbf{E}(X_i) = \frac{1}{n!}\sum_{j=1}^{n} a(i,j)\,\mathrm{card}\{\pi \mid \pi(i) = j\} = \frac{1}{n}\sum_{j=1}^{n} a(i,j) =: p_i$$
$$\mathbf{E}(X) = \sum_{i=1}^{n} p_i = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a(i,j) =: \lambda$$

zato je porazdelitev slučajne spremenljivke X smiselno aproksimirati s Poissonovo porazdelitvijo $\mathrm{Po}(\lambda)$.

Oprli se bomo na izrek 2.2.3. Pri dokazovanju pogojev tega izreka bomo začeli tako kot v krajšem dokazu leme 2.2.1. V njem moramo med drugim izračunati tudi $\mathbf{E}(X_i h(X))$. Velja:

$$\mathbf{E}\big(X_i h(X)\big) = \frac{1}{n}\sum_{j=1}^{n} a(i,j)\,\mathbf{E}\big(h(X) \mid A_{ij}\big)$$

Označili smo $A_{ij} := \{\pi \mid \pi(i) = j\}$ in upoštevali, da je $\mathbf{P}(A_{ij}) = \frac{1}{n}$.

V naslednjem koraku bomo pogojno porazdelitev glede na dogodek A_{ij} izrazili z brezpogojno porazdelitvijo. To bomo storili tako, da bomo konstruirali projekcijo $p_{ij}\colon \Omega \to A_{ij}$, za katero bo veljalo, da se mera $p_{ij*}\mathbf{P}$ ujema z normalizirano zožitvijo mere $\mathbf{P}$ na A_{ij}. Natančneje, veljati mora:

$$\mathbf{P}\big(p_{ij}^{-1}(B)\big) = \frac{\mathbf{P}(B)}{\mathbf{P}(A_{ij})} = n\mathbf{P}(B)$$

Očitno je to dovolj preveriti na točkah. Za vsako permutacijo $\pi \in A_{ij}$ bo moralo torej vlakno $p_{ij}^{-1}(\{\pi\})$ vsebovati natanko n točk. V tem primeru bo pogojna porazdelitev slučajne spremenljivke X glede na dogodek A_{ij} enaka brezpogojni porazdelitvi slučajne spremenljivke $X \circ p_{ij}$. *Na istem verjetnostnem prostoru* smo torej definirali slučajni spremenljivki, od katerih ima ena (namreč X) enako porazdelitev kot X, porazdelitev druge pa se ujema s pogojno porazdelitvijo slučajne spremenljivke X glede na dogodek A_{ij}. Temu pravimo *sklapljanje*. Sklapljanje je eden glavnih prijemov pri Chen–Steinovi metodi in ima zelo široko področje uporabe. Tako cela knjiga [6], ki obravnava Poissonovo aproksimacijo po Chen–Steinovi metodi, temelji le na sklapljanju.

Definirajmo projekcijo p_{ij}: $S \to A_{ij}$ po predpisu:

$$p_{ij}(\rho)(k) := \begin{cases} j & k = i \\ \rho(i) & k = \rho^{-1}(j) \\ \rho(k) & \text{sicer} \end{cases}$$

V primeru, ko je $\rho(i) = j$ oz. $\rho^{-1}(j) = i$, prvi dve vrstici sovpadata in velja $p_{ij}(\rho) = \rho$. Očitno p_{ij} res slika v A_{ij}. Naj bo zdaj $\pi \in A_{ij}$. Preveriti moramo, da obstaja natanko n permutacij ρ, za katere je $p_{ij}(\rho) = \pi$.

Za vsak $k = 1, \ldots n$ poglejmo, koliko je permutacij ρ, za katere je $\rho^{-1}(j) = k$ in $p_{ij}(\rho) = \pi$. Očitno mora ρ na vseh elementih, ki so različni od i in k, delovati tako kot π. Biti mora še $\rho(k) = j$, s tem pa je ρ natančno določena. Očitno za tako določeno permutacijo ρ res velja $p_{ij}(\rho) = \pi$, torej ima $p_{ij}^{-1}(\{\pi\})$ res natanko n elementov.

Velja torej $\mathcal{L}(X \mid A_{ij}) = \mathcal{L}(X \circ p_{ij})$. Sledi:

$$\mathbf{E}\big(X_i h(X)\big) = \frac{1}{n} \sum_{j=1}^{n} a(i,j) \mathbf{E}\big(h(X \circ p_{ij})\big)$$

Za vsako permutacijo π velja:

$$X(p_{ij}(\pi)) = X(\pi) - a(i, \pi(i)) - a(\pi^{-1}(j), j) + a(i,j) + a(\pi^{-1}(j), \pi(i))$$

Napišimo to bolj kompaktno. Označimo:

$$\begin{aligned} Y_j(\pi) &:= a(\pi^{-1}(j), j) \\ Z_{ij}(\pi) &:= a(\pi^{-1}(j), \pi(i)) \end{aligned}$$

Tedaj velja:

$$X \circ p_{ij} = X - X_i - Y_j + a(i,j) + Z_{ij}$$

Vse skupaj bomo morali pomnožiti z $a(i,j)$. Če je $a(i,j) = 0$, bomo dobili 0, če je $a(i,j) = 1$, pa velja:

$$\begin{aligned} X \circ p_{ij} &= X + 1 + Z_{ij} - X_i - Y_j \\ h(X \circ p_{ij}) &= h(X+1) + I_{[Z_{ij}=1]}(\Delta h)(X+2) - I_{[X_i=1]}(\Delta h)(X+1+Z_{ij}) - \\ &\quad - I_{[Y_j=1]}(\Delta h)(X+1+Z_{ij}-X_i) \end{aligned}$$

Če definiramo:

$$R_{ij} := \mathbf{E}\big(h(X+1) - h(X \circ p_{ij})\big)$$

velja:

$$\mathbf{E}\big(h(X \circ p_{ij})\big) = \mathbf{E}\big(h(X+1)\big) - R_{ij}$$

Če bomo hoteli zadostiti pogojem izreka 2.2.3, bomo morali argumente funkcije Δh oceniti z $M_1(h) = \sup_{r \geq 2} \big|(\Delta h)(r)\big|$. Preveriti moramo torej, da so vsi argumenti v Δh enaki vsaj 2, brž ko je faktor pred Δh različen od 0. Za argument $X + 2$ to že velja. Če je $X_i = 1$, je tudi $X = \sum_{k=1}^{n} X_k \geq 1$, torej je tudi $X + 1 + Z_{ij} \geq 2$. Preverimo še tretji argument. Naj bo $Y_j = 1$. Velja $Y_j(\pi) = a(\pi^{-1}(j), j) = X_{\pi^{-1}(j)}(\pi)$. Brž ko je $Y_j(\pi) = 1$, je torej tudi $X(\pi) \geq 1$. Edina možnost, pri kateri bi bil lahko argument v Δh manjši od 2, je, da je še $X_i(\pi) = 1$. Če je $\pi(i) \neq j$, je $X(\pi) \geq X_i(\pi) + X_{\pi^{-1}(j)}(\pi) = 2$ in zadeva velja. Če pa je $\pi(i) = j$, je $Z_{ij}(\pi) = X_i(\pi)$ in ker je $X(\pi) \geq 1$, zadeva tudi drži. Torej lahko ocenimo:

$$|R_{ij}| \leq M_1(h)\mathbf{E}\big(X_i + Y_j + Z_{ij}\big)$$

Velja:

$$\mathbf{E}(X_i) = \frac{1}{n}\sum_{j=1}^{n} a(i,j) = p_i$$

$$\mathbf{E}(Y_j) = \frac{1}{n}\sum_{i=1}^{n} a(i,j) =: q_j$$

Izračunati moramo še:

$$\mathbf{E}(Z_{ij}) = \sum_{r=1}^{n}\sum_{s=1}^{n} a(r,s)\mathbf{P}(B_{ijrs})$$

kjer je:

$$B_{ijrs} := \{\pi \mid \pi^{-1}(j) = r, \pi(i) = s\}$$

Če je $r = i$ in želimo, da množica B_{ijrs} ni prazna, mora biti očitno tudi $s = j$. V tem primeru je $\mathbf{P}(B_{ijrs}) = \frac{1}{n}$. Če pa je $r \neq i$ in želimo, da B_{ijrs} ni prazna, mora biti očitno tudi $s \neq j$. V tem primeru pa ima permutacija π predpisani dve vrednosti, torej je $\mathbf{P}(B_{ijrs}) = \frac{1}{n(n-1)}$. Sledi:

$$\begin{aligned}
\mathbf{E}(Z_{ij}) &= \frac{1}{n}a(i,j) + \frac{1}{n(n-1)}\sum_{r\neq i}\sum_{s\neq j} a(r,s) = \\
&= \frac{1}{n}a(i,j) + \frac{1}{n(n-1)}\left(\sum_{r=1}^{n}\sum_{s=1}^{n} a(r,s) - \sum_{s=1}^{n} a(i,s) - \sum_{r=1}^{n} a(r,j) + a(i,j)\right) = \\
&= \frac{1}{n-1}\left(a(i,j) + \lambda - p_i - q_j\right) \\
|R_{ij}| &\leq \frac{M_1(h)}{n-1}\big(a(i,j) + \lambda + (n-2)(p_i + q_j)\big)
\end{aligned}$$

Lotimo se zdaj končno ocenjevanja po Chen–Steinovi metodi. Velja:

$$\begin{aligned}
\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big) &= \lambda\mathbf{E}\big(h(X+1)\big) - \sum_{i=1}^{n} X_i h(X) = \\
&= \lambda\mathbf{E}\big(h(X+1)\big) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a(i,j)\mathbf{E}\big(h(X \circ p_{ij})\big) = \\
&= \lambda\mathbf{E}\big(h(X+1)\big) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a(i,j)\Big(\mathbf{E}\big(h(X+1)\big) - R_{ij}\Big) = \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a(i,j)R_{ij}
\end{aligned}$$

Sledi:

$$\begin{aligned}\left|\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big)\right| &\leq \frac{M_1(h)}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} a(i,j)\big(a(i,j) + \lambda + (n-2)(p_i+q_j)\big) = \\ &= \frac{M_1(h)}{n-1}\left(\lambda + \lambda^2 + (n-2)\left(\sum_{i=1}^{n} p_i^2 + \sum_{j=1}^{n} q_j^2\right)\right)\end{aligned}$$

Iz izreka 2.2.3 zdaj končno sledi:

$$\begin{aligned}d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) &\leq \frac{1-e^{-\lambda}}{(n-1)\lambda}\left(\lambda + \lambda^2 + (n-2)\left(\sum_{i=1}^{n} p_i^2 + \sum_{j=1}^{n} q_j^2\right)\right) \leq \\ &\leq \frac{\min\left\{1, \frac{1}{\lambda}\right\}}{n-1}\left(\lambda + \lambda^2 + (n-2)\left(\sum_{i=1}^{n} p_i^2 + \sum_{j=1}^{n} q_j^2\right)\right)\end{aligned}$$

Spet smo dobili lepo, pregledno oceno, ki spominja na oceno iz izreka 2.2.4. V primeru, ko so vsi parametri p_i in q_j enaki, torej v primeru, ko je $p_i = q_j = \frac{\lambda}{n}$, dobimo:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \leq \frac{1-e^{-\lambda}}{n-1}\left(1 + \frac{3n-4}{n}\lambda\right)$$

Hitrost konvergence je torej istega velikostnega reda, kot če bi bile slučajne spremenljivke $X_1, \ldots X_n$ neodvisne, namreč reda $\frac{1}{n}$. Chen–Steinova metoda nam je torej s sicer nekaj več računanja za odvisne slučajne spremenljivke dala prav tako kakovosten rezultat.

2.3 Lokalna odvisnost

V drugem zgledu v prejšnjem razdelku smo videli, da Chen–Steinova metoda lepo deluje tudi za odvisne slučajne spremenljivke. Pogoje izreka 2.2.3 smo seveda morali preveriti na roko, saj nam je zaenkrat na voljo le lema 2.2.1, ki obravnava le neodvisne slučajne spremenljivke. V tem razdelku jo bomo posplošili na slučajne spremenljivke z dano strukturo *lokalne odvisnosti*. Tudi slučajnih spremenljivk ne bo nujno več končno mnogo, ampak bomo obravnavali teoretično poljubne, v praksi pa največ števno neskončne družine slučajnih spremenljivk X_α, kjer α pripada indeksni množici $\mathcal{I}$. Sledili bomo članku [2] in vsakemu indeksu α priredili *okolico odvisnosti* B_α, znotraj katere odvisnost slučajnih spremenljivk podrobno poznamo. Natančneje, za vsak $\beta \in B_\alpha$ bo treba poznati korelacijo med X_α in X_β. Zunaj okolice odvisnosti B_α pa bomo odvisnost le grobo ocenili.

Naj bodo torej X_α, kjer je $\alpha \in \mathcal{I}$, Bernoullijeve slučajne spremenljivke in naj bo $p_\alpha := \mathbf{P}[X_\alpha = 1]$. Naj bo $X = \sum_{\alpha\in\mathcal{I}} X_\alpha$. Ta slučajna spremenljivka ima zdaj sicer vrednosti v $\mathbb{N}_0 \cup \{\infty\}$, vendar je skoraj gotovo končna, če je $\lambda := \sum_{\alpha\in\mathcal{I}} p_\alpha < \infty$. Za vsak indeks $\alpha \in \mathcal{I}$ izberimo množico $B_\alpha \subset \mathcal{I}$, ki vsebuje α. Označimo:

$$\begin{aligned}V_\alpha &:= \sum_{\beta\in\mathcal{I}\setminus B_\alpha} X_\beta \\ W_\alpha &:= X - X_\alpha\end{aligned}$$

Na podoben način kot v krajšem dokazu leme 2.2.1 bomo ocenili izraz $\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big)$. Tu pa h ne bo kar poljubna, temveč *omejena* funkcija, saj bi drugače imeli težave z dobro definiranostjo tega izraza, pa tudi v sami oceni bo nastopala tudi supremum norma funkcije h. To pomeni, da bo za oceno v metriki totalne variacije treba oceniti tudi normo same funkcije h in ne le Δh, kjer je h rešitev enačbe (2.2.9). Treba bo torej oceniti tudi *prve*, ne le druge diference rešitve Steinove enačbe. Pripomniti velja, da so izključno s prvimi diferencami izpeljane tudi precej ostre ocene v članku [25].

Naj bo torej $h\colon \mathbb{N}_0 \to \mathbb{R}$ omejena funkcija. Iz zvez:

$$\begin{aligned} X_\alpha h(X) &= X_\alpha h(W_\alpha + 1) \\ h(X+1) - h(W_\alpha + 1) &= X_\alpha (\Delta h)(W_\alpha + 2) \end{aligned}$$

podobno kot v krajšem dokazu leme 2.2.1 sledi:

$$\begin{aligned} &\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big) = \\ &\quad = \sum_{\alpha\in\mathcal{I}} \mathbf{E}\big(p_\alpha h(X+1) - X_\alpha h(X)\big) = \\ &\quad = \sum_{\alpha\in\mathcal{I}} \mathbf{E}\Big(p_\alpha h(X+1) - p_\alpha h(W_\alpha+1) + p_\alpha h(W_\alpha+1) - X_\alpha h(W_\alpha+1)\Big) = \\ &\quad = \sum_{\alpha\in\mathcal{I}} \mathbf{E}\Big(p_\alpha X_\alpha (\Delta h)(W_\alpha+2)\big) + (p_\alpha - X_\alpha)\big(h(W_\alpha+1) - h(V_\alpha+1)\big) + \\ &\qquad + (p_\alpha - X_\alpha) h(V_\alpha+1)\Big) \end{aligned}$$

Očitno je:

$$\big|\mathbf{E}\big(p_\alpha X_\alpha (\Delta h)(W_\alpha+2)\big)\big| \le M_1(h)\, p_\alpha^2$$

kjer je $M_1(h) = \sup_{r\ge 2}\big|(\Delta h)(r)\big|$. Očitno velja tudi tole: če je $1 \le n \le m$, je $|h(m) - h(n)| \le M_1(h)(m-n)$. Sledi:

$$\begin{aligned} \Big|\mathbf{E}\Big((p_\alpha - X_\alpha)\big(h(W_\alpha+1) - h(V_\alpha+1)\big)\Big)\Big| &\le \\ \le M_1(h)\, \mathbf{E}\left(\big|(p_\alpha - X_\alpha)(W_\alpha - V_\alpha)\big|\right) &= \\ = M_1(h)\, \mathbf{E}\Big(\Big|(p_\alpha - X_\alpha){\textstyle\sum_{\beta\in B_\alpha\setminus\{\alpha\}}} X_\alpha\Big|\Big) &\le \\ \le M_1(h) \sum_{\beta\in B_\alpha\setminus\{\alpha\}} (p_{\alpha\beta} + p_\alpha p_\beta) \end{aligned}$$

kjer je:

$$p_{\alpha\beta} := E(X_\alpha X_\beta)$$

Oceniti je treba še zadnji sumand. Velja:

$$\begin{aligned} \big|\mathbf{E}\big((p_\alpha - X_\alpha) h(V_\alpha+1)\big)\big| &= \mathbf{E}\left(\Big|\mathbf{E}\left((p_\alpha - X_\alpha) h(V_\alpha+1) \,\Big|\, {\textstyle\sum_{\beta\in\mathcal{I}\setminus B_\alpha}} X_\beta\right)\Big|\right) \\ &\le M_0(h)\, \mathbf{E}\left(\Big|\mathbf{E}\left(p_\alpha - X_\alpha \,\Big|\, {\textstyle\sum_{\beta\in\mathcal{I}\setminus B_\alpha}} X_\beta\right)\Big|\right) \end{aligned}$$

kjer je:

$$M_0(h) := \sup_{r\geq 1}|h(r)|$$

Zgornje ocene lahko strnemo v naslednjo lemo.

Lema 2.3.1 (Arratia, Goldstein, Gordon). *Naj bodo X_α, $\alpha \in \mathcal{I}$, Bernoullijeve slučajne spremenljivke. Naj bo $X := \sum_{\alpha\in\mathcal{I}} X_\alpha$ in $\lambda := \mathbf{E}(X) < \infty$. Za vsak $\alpha \in \mathcal{I}$ izberimo množico $B_\alpha \subset \mathcal{I}$, ki vsebuje α. Definirajmo:*

$$\begin{aligned}
p_\alpha &:= \mathbf{E}(X_\alpha), \quad p_{\alpha\beta} := \mathbf{E}(X_\alpha X_\beta)\\
b_1 &:= \sum_{\alpha\in\mathcal{I}}\sum_{\beta\in B_\alpha} p_\alpha p_\beta\\
b_2 &:= \sum_{\alpha\in\mathcal{I}}\sum_{\beta\in B_\alpha\setminus\{\alpha\}} p_{\alpha\beta}\\
b_3 &:= \sum_{\alpha\in\mathcal{I}} \mathbf{E}\left(\left|\mathbf{E}\left(p_\alpha - X_\alpha \,\middle|\, \textstyle\sum_{\beta\in\mathcal{I}\setminus B_\alpha} X_\beta\right)\right|\right) \leq\\
&\leq \sum_{\alpha\in\mathcal{I}} \mathbf{E}\left(\left|\mathbf{E}\left(p_\alpha - X_\alpha \,\middle|\, \sigma(X_\beta \mid \beta\in\mathcal{I}\setminus B_\alpha)\right)\right|\right)
\end{aligned}$$

Tedaj za vsako omejeno funkcijo h: $\mathbb{N}_0 \to \mathbb{R}$ velja ocena:

$$\left|\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big)\right| \leq (b_1+b_2)\,M_1(h) + b_3\,M_0(h)$$

■

Opomba. Površno povedano, parameter b_1 meri število sosednih dogodkov, parameter b_2 meri število sosednih dogodkov, ki se zgodijo skupaj z osrednjim, parameter b_3 pa meri vpliv dogodkov izven okolice odvisnosti. Zgornja lema je posebej uporabna v primeru, ko je slučajna spremenljivka X_α popolnoma neodvisna od dogodkov izven okolice odvisnosti. V tem primeru je seveda $b_3 = 0$.

Lema 2.3.1 je seveda posplošitev leme 2.2.1 na lokalno odvisne slučajne spremenljivke. Seveda pa nam lema 2.2.1 še ni dala dejanske ocene napake pri Poissonovi aproksimaciji, oceniti je bilo treba še rešitev Steinove enačbe, kar smo storili v lemi 2.2.2. Le-ta nam je dala oceno drugih diferenc, torej oceno za $M_1(h)$. Tu pa druge diference ne bodo dovolj, oceniti bo treba še prve diference, torej $M_0(h)$.

Lema 2.3.2. *Naj bo $A \subset \mathbb{N}_0$ in naj bo h_A rešitev prirejene Steinove enačbe:*

$$\lambda h_A(r+1) - rh_A(r) = I_A(r) - \mathrm{Po}(\lambda)(A) \tag{2.3.1}$$

Tedaj velja ocena:

$$M_0(h_A) \leq \min\left\{1, \tfrac{5}{4}\lambda^{-1/2}\right\}$$

Opomba. Ocena v zgornji lemi je boljša od tiste, ki jo navaja članek [2] in je dokazana v članku [5]. V slednjem članku namreč namesto konstante $\frac{5}{4}$ stoji $\frac{7}{5}$.

Opomba. Dokazali bomo, da je konstanta 1 tudi najboljša možna ocena, ki velja za vse λ. Del ocene za velike λ pa se da še nekoliko izboljšati. V [6] je na strani 222 (opomba 10.2.4) dokazana naslednja ocena:

$$M_0(h_A) \leq \min\left\{1, \sqrt{\frac{2}{e\lambda}}\right\} \tag{2.3.2}$$

Lemo 2.3.2 bomo dokazali z elementarnimi sredstvi. Izboljšana ocena v [6] pa se dokaže s pomočjo procesa rojevanja in umiranja, ki tiči v ozadju Chen–Steinove metode (glej stran 137).

Opomba. Na oceni (2.3.2) temelji tudi članek [25], v katerem so izpeljane zelo ostre ocene napake pri Poissonovi aproksimaciji, med drugim tudi ocena (2.2.11).

Dokaz leme. Po formuli (2.2.3) za $r = 1, 2, \ldots$ velja:

$$h_A(r) = \frac{(r-1)!}{\lambda^r} \sum_{i=0}^{r-1} \frac{\lambda^i}{i!} \varphi_A(i) \tag{2.3.3}$$

kjer je $\varphi_A = I_A - \mathrm{Po}(\lambda)(A)$. Oglejmo si najprej $h_{\{j\}}$, kjer je $j \in \mathbb{N}_0$. Za $i \neq j$ je očitno $\varphi_{\{j\}}(i) \leq 0$, torej za $r \leq j$ velja $h_{\{j\}}(r) \leq 0$. Po formuli (2.2.5) je tudi:

$$h_A(r) = -\frac{(r-1)!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{i!} \varphi_A(i) \tag{2.3.4}$$

Od tod sledi, da za $r > j$ velja $h_{\{j\}}(r) \geq 0$. Za fiksen r je funkcija $A \mapsto h_A(r)$ mera. Iz prejšnjih ugotovitev potem sledi:

$$h_A(r) \leq h_{\{0,\ldots r-1\}}(r) \tag{2.3.5}$$

Za $r = 1$ dobimo:

$$h_A(1) \leq h_{\{0\}}(1) = \frac{\varphi_{\{0\}}(0)}{\lambda} = \frac{1 - e^{-\lambda}}{\lambda} \leq 1 \tag{2.3.6}$$

Ta meja je tudi asimptotično dosežena pri $A = \{0\}$ in $\lambda \downarrow 0$, torej je konstanta 1 res najboljša možna ocena, ki je neodvisna od λ. Za $r = 2$ dobimo:

$$\begin{aligned} h_A(2) \leq h_{\{0,1\}}(2) = \frac{\varphi_{\{0,1\}}(0) + \lambda\varphi_{\{0,1\}}(1)}{\lambda^2} = \frac{(1+\lambda)(1 - e^{-\lambda} - \lambda e^{-\lambda})}{\lambda^2} = \\ = \frac{1 + \lambda - (1+\lambda)^2 e^{-\lambda}}{\lambda^2} \end{aligned} \tag{2.3.7}$$

Ta izraz moramo zdaj navzgor omejiti z 1. Definirajmo:

$$\rho(\lambda) := 1 + \lambda - (1+\lambda)^2 e^{-\lambda}$$

Dovolj bo dokazati, da je $\rho(\lambda) \leq \lambda^2$. Očitno je $\rho(0) = 0$. Če odvajamo, dobimo:

$$\begin{aligned} \rho'(\lambda) &= 1 - 2(1+\lambda)e^{-\lambda} + (1-\lambda)^2 e^{-\lambda}, \quad \rho'(0) = 0 \\ \rho''(\lambda) &= (1 + 2\lambda - \lambda^2)e^{-\lambda} \end{aligned}$$

Trdimo: $\rho''(\lambda) \le 2$. Za ta namen poiščimo ekstreme funkcije ρ''. Velja:

$$\rho'''(\lambda) = (\lambda^2 - 4\lambda + 1)e^{-\lambda} = \big((\lambda-2)^2 - 3\big)e^{-\lambda}$$

Velja $\rho''(0) = 1$. Nato funkcija ρ'' narašča do $\lambda = 2 - \sqrt{3}$, nato pada do $\lambda = 2 + \sqrt{3}$ in nazadnje spet narašča. Njena limita, ko gre λ proti neskončno, je enaka 0. Zato za vsak $\lambda \ge 0$ velja:

$$\rho''(\lambda) \le \rho''\big(2-\sqrt{3}\big) = 2\big(\sqrt{3}-1\big)e^{\sqrt{3}-2} < 2$$

Ko to dvakrat integriramo, dobimo, da res za vsak $\lambda > 0$ velja $\rho(\lambda) \le \lambda^2$, potem pa je tudi $h_A(2) \le 1$ za vsak A in vsak λ.

Oceno za $r > 2$ pa bomo naredili drugače. Naj bo najprej $\lambda \ge r - 1$. V tem primeru uporabimo formulo (2.3.3). Očitno je $\varphi_A \le 1$, torej velja:

$$\begin{aligned} h_A(r) &\le \frac{(r-1)!}{\lambda^r} \sum_{i=0}^{r-1} \frac{\lambda^i}{i!} = \sum_{i=0}^{r-1} \frac{(r-1)(r-2)\ldots(i+1)}{\lambda^{r-1}} \le \\ &\le \sum_{i=0}^{r-1} \frac{(r-1)^{r-i-1}}{\lambda^{r-i}} = \frac{1 - \left(\frac{r-1}{\lambda}\right)^r}{\lambda - r + 1} \end{aligned} \tag{2.3.8}$$

V primeru, ko je $\lambda \le r + 1$, pa uporabimo formulo (2.3.4) in upoštevamo, da je $\varphi_A \ge -1$. Dobimo:

$$\begin{aligned} h_A(r) &\le \frac{(r-1)!}{\lambda^r} \sum_{i=r}^{\infty} \frac{\lambda^i}{i!} = \sum_{i=r}^{\infty} \frac{\lambda^{i-r}}{r(r+1)\ldots i} \le \\ &\le \frac{1}{r} \sum_{i=r}^{\infty} \frac{\lambda^{i-r}}{(r+1)^{i-r}} = \frac{r+1}{r(r+1-\lambda)} \end{aligned} \tag{2.3.9}$$

V primeru, ko je $\lambda \le r - \frac{1}{r}$, nam ta ocena že da zahtevano oceno $h_A(r) \le 1$. V primeru, ko je $\lambda \ge r - \frac{1}{r}$, pa bomo uporabili oceno (2.3.8). Iz geometrijske vrste je jasno, da je zadnji izraz v oceni (2.3.8) padajoča funkcija parametra λ. Za $\lambda \ge r - \frac{1}{r}$ torej velja:

$$h_A(r) \le \frac{1 - \left(\frac{r-1}{r-\frac{1}{r}}\right)^r}{r - \frac{1}{r} - r + 1} = \frac{1 - \left(1 + \frac{1}{r}\right)^{-r}}{1 - \frac{1}{r}}$$

Oceno je dovolj narediti za $r \ge 3$. Za $r = 3$ dobimo:

$$h_A(3) \le \frac{1 - \left(\frac{3}{4}\right)^3}{\frac{2}{3}} = \frac{111}{128} < 1$$

za $r \ge 4$ pa dobimo:

$$h_A(r) \le \tfrac{4}{3}\left(1 - \left(1 + \tfrac{1}{r}\right)^{-r}\right) \le \tfrac{4}{3}\left(1 - \tfrac{1}{e}\right) \doteq 0.84 < 1$$

Znak $\doteq$ pomeni, da enakost velja na toliko decimalnih mest, kolikor jih ima decimalni ulomek na njegovi levi ali desni. Dokazali smo, da za vsak A, vsak $\lambda > 0$ in vsak $r \in \mathbb{N}$ velja ocena $h_A(r) \le 1$. Ker je $h_{A^c} = -h_A$, je tudi $h_A(r) \ge -1$, potem pa je res $M_0(h_A) \le 1$.

Dokazati moramo še oceno $M_0(h_A) \leq \frac{5}{4}\lambda^{-1/2}$. Za ta namen zapišimo rešitev enačbe (2.3.1) nekoliko drugače. Naj bo $U_r := \{0, \dots r-1\}$. Iz formule (2.3.3) sledi:

$$\begin{aligned} h_A(r) &= \frac{e^\lambda (r-1)!}{\lambda^r} \int_{U_r} \big(I_A - \mathrm{Po}(\lambda)(A)\big)\, d\,\mathrm{Po}(\lambda) = \\ &= \frac{e^\lambda (r-1)!}{\lambda^r} \Big(\mathrm{Po}(\lambda)(U_r \cap A) - \mathrm{Po}(\lambda)(U_r)\ \mathrm{Po}(\lambda)(A)\Big) \end{aligned}$$

Iz ocene (2.3.5) sledi:

$$h_A(r) \leq h_{U_r}(r) = \frac{e^\lambda (r-1)!}{\lambda^r} \Big(\mathrm{Po}(\lambda)(U_r) - \mathrm{Po}(\lambda)(U_r)^2\Big) \leq \frac{e^\lambda (r-1)!}{4\lambda^r}$$

Idejo za nadaljnje ocenjevanje si bomo izposodili pri Stirlingovem obrazcu. Pišimo:

$$n! = C_n \left(\frac{n}{e}\right)^e \sqrt{n}$$

Stirlingov obrazec nam pove, da je $\lim_{n\to\infty} = \sqrt{2\pi}$. Mi pa bi radi konstante C_n ocenili navzgor z vrednostjo, ki bi bila čim bliže $\sqrt{2\pi}$. Pokažimo, da je zaporedje C_n padajoče. Krajši račun pokaže, da je:

$$\frac{C_{n+1}}{C_n} = e \left(\frac{n}{n+1}\right)^{n+\frac{1}{2}}$$

Postavimo $a := n + \frac{1}{2}$. Tedaj velja:

$$\begin{aligned} \log \frac{C_{n+1}}{C_n} &= 1 + a \log\left(1 - \frac{1}{2a}\right) - a \log\left(1 + \frac{1}{2a}\right) = \\ &= 1 - 2a \left(\frac{1}{2a} + \frac{1}{3}\left(\frac{1}{2a}\right)^3 + \frac{1}{5}\left(\frac{1}{2a}\right)^5 + \dots\right) < 0 \end{aligned}$$

Za $n \geq 2$ je torej $C_n \leq C_2 = \frac{e^2\sqrt{2}}{4} \doteq 2.61$, medtem ko je $\sqrt{2\pi} \doteq 2.51$. Razlika ni prevelika. Če torej označimo $C := \frac{1}{4}C_2$, za $r \geq 3$ velja:

$$\begin{aligned} h_A(r) &\leq C\, \frac{e^\lambda}{\lambda^r} \left(\frac{r-1}{e}\right)^{r-1} \sqrt{r-1} = \frac{C}{\sqrt{\lambda}}\, e^{\lambda - r + 1} \left(\frac{r-1}{\lambda}\right)^{r-\frac{1}{2}} = \\ &= \frac{C}{\sqrt{\lambda}}\, e^{\lambda-r+1} \left(1 + \frac{r-\lambda-1}{\lambda}\right)^{r-\frac{1}{2}} \leq \frac{C}{\sqrt{\lambda}}\, e^{\lambda-r+1} e^{\left(r-\frac{1}{2}\right)\frac{r-\lambda-1}{\lambda}} = \\ &= \frac{C}{\sqrt{\lambda}}\, e^{\frac{(\lambda-r+1)(\lambda-r+\frac{1}{2})}{\lambda}} \leq \frac{C}{\sqrt{\lambda}}\, e^{\frac{(\lambda-r+\frac{3}{4})^2}{\lambda}} \end{aligned}$$

Naj bo $\alpha > 0$. Če je $\lambda - \sqrt{\lambda} + \frac{3}{4} \leq r \leq \lambda + \sqrt{\lambda} + \frac{3}{4}$, lahko ocenimo:

$$h_A(r) \leq \frac{e^2\sqrt{2}}{16\sqrt{\lambda}}\, e^{\alpha^2} \tag{2.3.10}$$

Če je $r \leq \lambda - \sqrt{\lambda} + \frac{3}{4}$, lahko po formuli (2.3.8) ocenimo:

$$h_A(r) \leq \frac{1 - \left(\frac{r-1}{\lambda}\right)^r}{\lambda - r + 1} \leq \frac{1}{\lambda - r + \frac{3}{4}} \leq \frac{1}{\alpha\sqrt{\lambda}} \tag{2.3.11}$$

Če pa je $r \geq \lambda + \sqrt{\lambda} + \frac{3}{4}$, pa uporabimo predzadnji izraz v verigi neenakosti (2.3.9). Dobimo:

$$h_A(r) \leq \frac{1}{r} \sum_{i=r}^{\infty} \frac{\lambda^{i-r}}{(r+1)^{i-r}} \leq \frac{1}{r} \sum_{i=0}^{\infty} \left(\frac{\lambda}{r}\right)^i = \frac{1}{r - \lambda} \leq \frac{1}{\alpha\sqrt{\lambda}} \tag{2.3.12}$$

Parameter α moramo izbrati tako, da so ocene (2.3.10), (2.3.11) in (2.3.12) čim boljše. To pa bo takrat, ko se bodo njihove desne strani ujemale. Nedaleč od ujemanja je izbira $\alpha := \frac{4}{5}$. Iz ocen (2.3.11) in (2.3.12) potem takoj sledi zahtevana ocena $h_A(r) \leq \frac{5}{4}\lambda^{-1/2}$, seveda za tiste r in λ, za katere oceni držita. Iz ocene (2.3.10) pa dobimo:

$$h_A(r) \leq \frac{e^{2.64}\sqrt{2}}{16\sqrt{\lambda}} \doteq 1.24\lambda^{-1/2} < \frac{5}{4}\lambda^{-1/2}$$

Dokazali smo, da za vse λ in vse $r \geq 3$ velja zahtevana ocena $h_A(r) \leq \frac{5}{4}\lambda^{-1/2}$. To moramo zdaj dokazati še za $r = 1$ in $r = 2$. Za $\lambda \leq \frac{25}{16}$ velja $h_A(r) \leq 1 \leq \frac{5}{4}\lambda^{-1/2}$. Naj bo $\lambda \geq \frac{25}{16}$. Za $r = 1$ po formuli (2.3.6) velja:

$$h_A(1) \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \frac{1}{\lambda} \leq \frac{5}{4}\lambda^{-1/2}$$

Za $r = 2$ pa iz ocene (2.3.7) dobimo:

$$h_A(2) \leq \frac{1 + \lambda - (1+\lambda)^2 e^{-\lambda}}{\lambda^2}$$

Dokazati moramo, da za $\lambda \geq \frac{25}{16}$ velja:

$$1 + \lambda - (1+\lambda)^2 e^{-\lambda} \leq \frac{5}{4}\lambda^{3/2}$$

Za $\frac{25}{16} \leq \lambda \leq 2$ velja:

$$1 + \lambda - (1+\lambda)^2 e^{-\lambda} \leq 3 - \left(1 + \frac{25}{16}\right)^2 e^{-2} \doteq 2.11 < 2.41 \doteq \frac{5}{4}\left(\frac{25}{16}\right)^{3/2} \leq \frac{5}{4}\lambda^{3/2}$$

za $\lambda \geq 2$ pa velja:

$$\begin{aligned} 1 + \lambda - (1+\lambda)^2 e^{-\lambda} &\leq 1 + \lambda = 3 + \int_2^\lambda dt \leq 3 + \int_2^\lambda \frac{15}{8}\sqrt{t}\,dt = \\ &= \frac{5}{4}\lambda^{3/2} - \frac{5}{4}2^{3/2} + 3 < \frac{5}{4}\lambda^{3/2} \end{aligned}$$

To pa pomeni, da za vse $\lambda \geq \frac{25}{16}$ velja $h_A(2) \leq \frac{5}{4}\lambda^{-1/2}$. Torej res za vse $r \in \mathbb{N}$ velja $h_A(r) \leq \frac{5}{4}\lambda^{-1/2}$. To pa je bilo treba dokazati. ∎

Opomba. V zgornjem dokazu bi lahko zgornjo mejo $\frac{5}{4}\lambda^{-1/2}$ še nekoliko izboljšali. Vendar pa bi imel tako spremenjen dokaz kar nekaj več tehničnih podrobnosti, zgornje meje pa ne bi prav bistveno izboljšali. Natančneje, v zgornjem dokazu smo morali oceno posebej izpeljati za $r = 1$ in $r = 2$, tako pa bi jo morali še za kak večji r (in po metodi, s katero smo dokazovali, je z vsakim naslednjim več dela).

Iz leme, ki smo jo pravkar dokazali, in leme 2.3.1 zdaj sledi naslednji izrek, ki je posplošitev izreka 2.2.3.

Izrek 2.3.3. *Naj bo X slučajna spremenljivka z vrednostmi v $\mathbb{N}_0$, naj bo $\lambda > 0$ in naj za vsako omejeno funkcijo $h\colon \mathbb{N}_0 \to \mathbb{R}$ velja:*

$$\mathbf{E}\big(\lambda h(X+1) - Xh(X)\big) \le c_0 M_0(h) + c_1 M_1(h)$$

Tedaj velja:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \le c_0 \min\left\{1, \tfrac{5}{4}\right\} \lambda^{-1/2} + c_1 \frac{1-e^{-\lambda}}{\lambda}$$

Posebej, naj bodo slučajne spremenljivke X_α, $\alpha \in \mathcal{I}$, okolice B_α ter λ, b_1, b_2 in b_3 kot v lemi 2.3.1. Tedaj velja:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \le (b_1 + b_2)\frac{1-e^{-\lambda}}{\lambda} + b_3 \min\left\{1, \tfrac{5}{4}\right\} \lambda^{-1/2}$$

2.4 Zgled: problem rojstnih dni

Klasični problem rojstnih dni se glasi takole: kolikšna je verjetnost, da imata vsaj dva od danih n ljudi rojstni dan na isti dan. Ta problem ni težak, saj se da brez težav napisati eksplicitna formula za verjetnost tega dogodka. Če nekoliko posplošimo, dobimo, da za poljubne neodvisne slučajne spremenljivke $Y_1, \ldots Y_n$, ki so enakomerno porazdeljene na množici z d elementi, velja:

$$\mathbf{P}[Y_1, \ldots Y_n \text{ so vse različne}] = \prod_{i=1}^{n-1}\left(1 - \frac{i}{d}\right) \tag{2.4.1}$$

Težava nastopi, če problem posplošimo. Slučajne spremenljivke $Y_1, \ldots Y_n$ npr. lahko niso več nujno enakomerno porazdeljene. Vprašamo se lahko tudi, kako je porazdeljeno število parov (i, j), za katere je $Y_i = Y_j$. Lahko tudi ne gledamo vseh možnih parov (i, j), ampak le vnaprej predpisane. Tovrstna posplošitev je obdelana v [6] na straneh 104–108, in sicer s pomočjo sklapljanja. Izsledki prejšnjega razdelka pa so bolj uporabni za drugačno posplošitev: zanimala nas bo porazdelitev števila podmnožic α moči k, za katere velja, da imajo vse slučajne spremenljivke Y_i, kjer je $i \in \alpha$, isto vrednost (za vrednosti Y_i, kjer je $i \notin \alpha$, pa se ne menimo: če so vse slučajne spremenljivke Y_i enake, je torej takih množic $\binom{n}{k}$). Zanimalo nas bo torej, na koliko načinov je mogoče izbrati k ljudi, ki vsi praznujejo rojstni dan na isti dan. Število teh načinov označimo z X. To je zdaj slučajna spremenljivka. Njeno porazdelitev bomo aproksimirali s Poissonovo in ocenili napako. Zgled povzemamo iz članka [1].

Najprej X zapišimo kot vsoto Bernoullijevih slučajnih spremenljivk. Indeksna množica naj bo:

$$\mathcal{I} := \big\{\alpha \subset \{1, \dots n\} \;\big|\; \operatorname{card} \alpha = k\big\}$$

Za vsak $\alpha \in \mathcal{I}$ naj bo X_α indikator dogodka, da so vse slučajne spremenljivke Y_i, kjer je $i \in \alpha$, enake. To je torej indikator dogodka, da vsi ljudje iz α praznujejo rojstni dan na isti dan. Potem je očitno $X = \sum_{\alpha \in \mathcal{I}} X_\alpha$. Slučajne spremenljivke X_α so seveda odvisne. Pač pa je, površno povedano, slučajna spremenljivka X_α neodvisna od vsega, kar se dogaja izven α. Če torej postavimo:

$$B_\alpha := \{\beta \in \mathcal{I} \mid \alpha \cap \beta \neq \emptyset\}$$

je X_α neodvisna od $\sigma(X_\beta \mid \beta \notin B_\alpha)$, zato je $b_3 = 0$. Očitno je:

$$\begin{aligned} p_\alpha &:= \mathbf{E}(X_\alpha) = d^{1-k} \\ \lambda &:= \mathbf{E}(X) = \binom{n}{k} d^{1-k} \end{aligned} \tag{2.4.2}$$

Porazdelitev slučajne spremenljivke X bomo torej aproksimirali s $\mathrm{Po}(\lambda)$. Preden se lotimo dela, pa si oglejmo še primer, ki bo ilustriral, kako natančna je Poissonova aproksimacija. Naj bo $d = 365$ in $k = 2$. Zanima nas verjetnost dogodka, da so vsi rojstni dnevi različni, torej $\mathbf{P}[X = 0]$. Ogledali si bomo primer, ko je ta verjetnost blizu polovici. Iz formule (2.4.1) dobimo:

$$\begin{aligned} n = 22: &\quad \mathbf{P}[X = 0] = 0.5243 \\ n = 23: &\quad \mathbf{P}[X = 0] = 0.4927 \end{aligned}$$

Poissonova aproksimacija pa nam da $\mathbf{P}[X = 0] \approx e^{-\lambda}$, kjer je $\lambda = n(n-1)/730$. Dobimo:

$$\begin{aligned} n = 22: &\quad \mathbf{P}[X = 0] \approx 0.5311 \\ n = 23: &\quad \mathbf{P}[X = 0] \approx 0.5000 \end{aligned}$$

Do razlike torej pride na drugi decimalki. V članku [1] je omenjeno le, da je v primeru, ko je $n = 23$, parameter λ enak $\ln 2$ na štiri decimalke natančno. To je sicer res, vendar pa je tudi zavajajoče, saj se natančna verjetnost od ene polovice razlikuje za bistveno več kot 0.001. Poissonova aproksimacija torej le ni tako natančna, kot namiguje [1].

Ocenimo zdaj napako za splošni primer. Držimo se oznak iz leme 2.3.1. Najprej je:

$$b_1 = \sum_{\alpha \in \mathcal{I}} \sum_{\beta \in B_\alpha} p_\alpha p_\beta = \operatorname{card}(\mathcal{I}) \operatorname{card}(B_\alpha) = \binom{n}{k} \left(\binom{n}{k} - \binom{n-k}{k} \right) d^{2-2k} \tag{2.4.3}$$

Za izračun konstante b_2 moramo izračunati še $\mathbf{E}(X_\alpha X_\beta)$. Dovolj se je omejiti na primer, ko je $\alpha \cap \beta \neq \emptyset$. V tem primeru je to verjetnost dogodka, da imajo vsi ljudje iz $\alpha \cup \beta$ rojstni dan na isti dan. Naj ima $\alpha \cap \beta$ j elementov. Tedaj je $\mathbf{E}(X_\alpha X_\beta) = d^{1-2k+j}$. Vseh množic $\beta \in \mathcal{I}$, za katere je $\operatorname{card}(\alpha \cap \beta) = j$, pa je $\binom{k}{j}\binom{n-k}{k-j}$. Sledi:

$$b_2 = \sum_{\alpha \in \mathcal{I}} \sum_{\beta \in B_\alpha \setminus \{\alpha\}} \mathbf{E}(X_\alpha X_\beta) = \binom{n}{k} \sum_{j=1}^{k-1} \binom{k}{j} \binom{n-k}{k-j} d^{1-2k+j} \tag{2.4.4}$$

Zdaj, ko smo izračunali vse potrebne konstante, lahko po izreku 2.3.3 že ocenimo napako. Vendar pa nam zapletena formula za b_2 ne pokaže prav nazorno, kako velika je napaka. Zato si oglejmo, kako se zadeva obnaša, ko je n velik, λ pa ne tako velik. Število n bomo torej poslali proti neskončno, medtem ko bo λ ostal približno isti. Račun bomo naredili posebej za $k = 2$ in $k > 2$, saj se izkaže, da je sta pri $k = 2$ parametra b_1 in b_2 istega velikostnega reda glede na n, pri $k > 2$ pa je parameter b_1 zanemarljiv v primerjavi z b_2. Za $k = 2$ sta formuli še precej enostavni in bomo naredili eksaktno oceno. Za $k > 2$ pa si bomo ogledali le asimptotično oceno, ki bo veljala za velike n.

Pri $k = 2$ iz formul (2.4.2), (2.4.3) in (2.4.4) dobimo:

$$\begin{aligned} d &= \frac{n(n-1)}{2\lambda} \\ b_1 &= \frac{n(n-1)}{2}\left(\frac{n(n-1)}{2} - \frac{(n-2)(n-3)}{2}\right) d^{-2} = \frac{4n-6}{n(n-1)}\lambda^2 \\ b_2 &= \frac{n(n-1)}{2}\, 2(n-2)\, \frac{4\lambda^2}{n^2(n-1)^2} = \frac{4n-8}{n(n-1)}\lambda^2 \end{aligned}$$

Iz izreka 2.3.3 potem dobimo:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \le \frac{8n-14}{n(n-1)}\lambda(1-e^{-\lambda}) = O\big(\tfrac{1}{n}\big) \tag{2.4.5}$$

Omenili smo že, da je problem rojstnih dni za $k = 2$ obdelan tudi v [6]. Za primerjavo povejmo, da iz izreka 5.G(ii) na strani 105 sledi:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \le \frac{4n-6}{n(n-1)}\lambda(1-e^{-\lambda}) \tag{2.4.6}$$

To je kar dvakrat boljša ocena kot (2.4.5). Vendar pa bomo na koncu razdelka oceno (2.4.5) izboljšali.

Ocenimo napako še za $k > 2$. Najprej je:

$$d = \left(\frac{\lambda}{\binom{n}{k}}\right)^{\frac{1}{1-k}} = \left(\frac{1}{\lambda}\binom{n}{k}\right)^{\frac{1}{k-1}} \approx \left(\frac{n^k}{k!\lambda}\right)^{\frac{1}{k-1}}$$

Znak $\approx$ tu pomeni približno enako za velike n. Natančneje, zapis $a(n) \approx b(n)$ pomeni, da je $\lim_{n\to\infty} \frac{a(n)}{b(n)} = 1$. Ocenimo zdaj konstanti b_1 in b_2. Velja:

$$\begin{aligned} \binom{n}{k} - \binom{n-k}{k} &= \frac{n(n-1)\ldots(n-k+1) - (n-k)(n-k-1)\ldots(n-2k+1)}{k!} \approx \\ &\approx \frac{n^{k-1}}{k!}\left(-\sum_{i=0}^{k-1} i + \sum_{i=k}^{2k-1} i\right) = \frac{k}{(k-1)!}n^{k-1} \end{aligned}$$

Sledi:

$$b_1 = \lambda^2 \frac{\binom{n}{k} - \binom{n-k}{k}}{\binom{n}{k}} \approx \lambda^2 \frac{kn^{k-1}k!}{(k-1)!n^k} = \frac{\lambda^2 k^2}{n} = O(n^{-1})$$

Za konstanto b_2 pa dobimo:

$$b_2 = \binom{n}{k} \sum_{j=1}^{k-1} \binom{k}{j} \binom{n-k}{k-j} \left(\frac{\lambda}{\binom{n}{k}} \right)^{\frac{1-2k+j}{1-k}} \approx \sum_{j=1}^{n} C(k,j,\lambda) n^k n^{k-j} n^{k \frac{1-2k+j}{k-1}} =$$
$$= \sum_{j=1}^{k-1} C(k,j,\lambda) n^{\frac{j-k}{k-1}} \approx C(k,k-1,\lambda) n^{-\frac{1}{k-1}} = O\left(n^{-\frac{1}{k-1}}\right)$$

Faktor pred $n^{-\frac{1}{k-1}}$ pa je enak:

$$C(k,k-1,\lambda) = \frac{k}{k!} (k!\lambda)^{\frac{1-2k+k-1}{1-k}} = k(k!)^{\frac{1}{k-1}} \lambda^{\frac{k}{k-1}}$$

Dobimo:

$$b_2 \approx k(k!)^{\frac{1}{k-1}} \lambda^{\frac{k}{k-1}} n^{-\frac{1}{k-1}}$$

Ker je $k > 2$, je konstanta b_1 zanemarljiva v primerjavi z b_2, ko je n velik. Dobimo torej:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \approx k(k!)^{-\frac{k-2}{k-1}} \lambda^{\frac{k}{k-1}} n^{-\frac{1}{k-1}} \frac{1-e^{-\lambda}}{\lambda} \leq$$
$$\leq k(k!)^{-\frac{k-2}{k-1}} n^{-\frac{1}{k-1}} \min\left\{\lambda^{\frac{1}{k-1}}, \lambda^{\frac{k}{k-1}}\right\}$$

Če povzamemo: ko gre n proti neskončno, λ pa je približno fiksen (nujna odstopanja zaradi celoštevilskosti gredo proti 0), porazdelitev števila skupin s skupnimi rojstnimi dnevi konvergira proti $\mathrm{Po}(\lambda)$, hitrost konvergence pa je reda $n^{-\frac{1}{k-1}}$.

Čeprav smo se ves čas ukvarjali le s primerom, ko so rojstni dnevi enakomerno porazdeljeni, se da na ta način napako oceniti tudi v primeru, ko se porazdelitev nekoliko razlikuje od enakomerne. Za $r = 1, \dots d$ naj bo $\pi_r := \mathbf{P}[Y_i = r]$ za vse i. Vsak posameznik naj ima torej z verjetnostjo p_r rojstni dan na dan r. Če je porazdelitev enakomerna, je seveda $\pi_r = \frac{1}{d}$ za vse r. Označimo:

$$\gamma := d \max_{1 \leq r \leq d} \pi_r$$

Tedaj je $\gamma \geq 1$ in porazdelitev je enakomerna natanko tedaj, ko je $\gamma = 1$. Če γ ni prevelik, lahko z zgornjim postopkom še vedno dobimo kar dobre ocene napake, saj se parametri p_α pomnožijo z največ γ^k, parametri $p_{\alpha\beta}$ pa z največ γ^{2k-1}. Parameter b_1 se torej pomnoži z največ γ^{2k}, parameter b_2 pa z največ γ^{2k-1}. Celotna ocena v metriki totalne variacije se nam torej pomnoži z največ γ^{2k}. Če je $k > 2$, pa se nam asimptotična ocena, ko gre n proti neskončno, pomnoži z največ γ^{2k-1}. Tako je ta metoda dobra tudi v primeru, če pri dnevih v letu upoštevamo še prestopna leta. V tem primeru je $\gamma = 366/365.25 \doteq 1.002$.

Omenili smo že, da je ocena napake pri $k = 2$ v [6] kar dvakrat boljša. Toda v primeru, ko so rojstni dnevi enakomerno porazdeljeni, se da okolica odvisnosti še nekoliko zmanjšati. Če namreč izberemo element $i \in \alpha$, je slučajna spremenljivka X_α neodvisna tudi od Y_i, torej od rojstnega dne fiksiranega posameznika i. Če izberemo npr. $i := \min(\alpha)$, lahko predpišemo:

$$B_\alpha := \{\beta \mid (\alpha \setminus \{\min(\alpha)\}) \cap \beta \neq \emptyset\}$$

Vendar pa se izkaže, da je ta izboljšava bistvena le pri $k = 2$ ali majhnem n. Pri $k > 2$ in velikem n je namreč parameter b_1 zanemarljiv, pri parametru b_2 pa bistven delež prispevajo le tisti parametri $p_{\alpha\beta}$, pri katerih ima presek $\alpha \cap \beta$ $k-1$ elementov. Če je $k > 2$, v tem primeru β avtomatično pripada tudi novi okolici B_α, torej se bistveni parametri ne spremenijo.

Pri $k = 2$ pa se nam ocena napake precej popravi. Nova okolica B_α ima le $n-1$ elementov, torej je:

$$b_1 = \binom{n}{2}(n-1)d^{-2} = \frac{2n-2}{n^2-n}\lambda^2$$

$$b_2 = \binom{n}{2}(n-2)d^{-2} = \frac{2n-4}{n^2-n}\lambda^2$$

Iz izreka 2.3.3 potem sledi:

$$d_{\mathrm{TV}}\big(\mathcal{L}(X), \mathrm{Po}(\lambda)\big) \leq \frac{4n-6}{n(n-1)}\lambda(1-e^{-\lambda}) \tag{2.4.7}$$

kar je natanko ocena (2.4.6), ki sledi iz izreka 5.G(ii) v [6]. Metoda lokalne neodvisnosti nam je torej dala natančno isto oceno kot sklapljanje. Krajši račun pa pokaže tudi, da se ocena napake (2.4.5), ki se da brez težav posplošiti na neenakomerno porazdelitev rojstnih dni, popolnoma ujema z oceno iz izreka 5.G(i) v [6], ki prav tako ocenjuje napako v primeru, ko porazdelitev rojstnih dni ni nujno enakomerna.

Videli smo, da je pri $k = 2$ ocena napake v primeru enakomerne porazdelitve približno dvakrat boljša od ocene v splošnem primeru. Toda manjše okolice B_α lahko gledamo tudi v primeru, ko rojstni dnevi niso enakomerno porazdeljeni, le da parameter b_3 potem ni več enak nič. Toda če je porazdelitev rojstnih dni dovolj blizu enakomerni (npr. pri dnevih v letu, kjer upoštevamo tudi prestopna leta), se da parameter b_3 dovolj dobro oceniti in tako še vedno dobimo oceno, ki je blizu oceni (2.4.7). Tovrstne izboljšave so možne tudi pri sklapljanju (glej izrek 5.H v [6]).

3.

Normalna aproksimacija

3.1 Motivacija

Naj bodo $X_1, \ldots X_n$ neodvisne in enako porazdeljene slučajne spremenljivke s končnim tretjim momentom. Brez škode za splošnost lahko privzamemo, da je $\mathbf{E}(X_k) = 0$ in $\mathrm{var}(X_k) = 1$. Označimo še $\alpha := \mathbf{E}(|X|)$ in $\beta := \mathbf{E}(|X|^3)$. Iz Jensenove neenakosti dobimo:

$$\alpha^2 \leq \mathbf{E}(|X|^2) = \mathrm{var}(X) = 1 = \big(\mathbf{E}(|X|^2)\big)^{3/2} \leq \mathbf{E}\big((|X|^2)^{3/2}\big) = \beta \tag{3.1.1}$$

torej je $\alpha \leq 1 \leq \beta$.

Iz centralnega limitnega izreka vemo, da porazdelitev slučajne spremenljivke:

$$X := \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$$

šibko konvergira proti standardni normalni porazdelitvi, ko gre n proti neskončno. Naš cilj pa je ne le pokazati šibko konvergenco, temveč tudi oceniti napako. Le-to bomo najprej ocenili v Wasserteinovi metriki (glej dodatek A, stran 76). Oceno povzemamo iz [3].

Začeli bomo natanko tako kot pri Poissonovi aproksimaciji. Izberimo torej neodvisne slučajne spremenljivke K in $Y_1, \ldots Y_n$ takole: slučajna spremenljivka K naj bo porazdeljena enakomerno na $\{1, \ldots n\}$ in za vsak k naj bo Y_k porazdeljena tako kot X_k. Slučajne spremenljivke K, $Y_1, \ldots Y_n$ naj bodo neodvisne in tudi neodvisne od $X_1, \ldots X_n$. Definirajmo $D_k := n^{-1/2}(Y_k - X_k)$ in naj bo $X' := X + D_K$. Tako slučajna spremenljivka X' ni preveč različna od X, če je le n dovolj velik. Podobno kot pri Poissonovi aproksimaciji lahko tudi tu preverimo, da je X' enako porazdeljena kot X (glej (2.1.1)).

Naj bo $g\colon \mathbb{R} \to \mathbb{R}$ dvakrat zvezno odvedljiva funkcija in za $i = 1, 2, 3$ naj bo:

$$M_i(g) := \sup_{x \neq y} \left| \frac{g^{(i-1)}(x) - g^{(i-1)}(y)}{x - y} \right| < \infty$$

Opomba. Količine M_i lahko izrazimo tudi kot supremume odvodov, saj velja naslednja trditev.

Trditev 3.1.1. *Naj bo funkcija $f\colon \mathbb{R} \to \mathbb{R}$ absolutno zvezna in integral svojega odvoda. Tedaj velja:*

$$\sup_{x \neq y} \left| \frac{f(x) - f(y)}{x - y} \right| = \operatorname*{ess\,sup}_x |f'(x)|$$

Opomba. Pogoj trditve je avtomatično izpolnjen, če je leva stran končna. Ni pa nujno izpolnjen, če je končna le desna stran: odvod funkcije je lahko povsod tam, kjer obstaja, enak 0, funkcija f pa ni konstantna.

Dokaz trditve. Označimo $a := \sup_{x \neq y} \left| \frac{f(x)-f(y)}{x-y} \right|$ in $b := \sup_x |f'(x)|$. Najprej velja:

$$|f(x) - f(y)| \leq \int_y^x |f'(t)|\, dt \leq b|x - y|$$

od koder sledi $a \leq b$. Pokažimo zdaj še nasprotno neenakost. Dovolj je pokazati za $b > 0$. Naj bo $0 < \varepsilon < \min\{1, b\}$. Obstaja taka merljiva množica $A \subset \mathbb{R}$, da je $m(A) > 0$, f odvedljiva v vseh točkah iz A in $|f'(x)| \geq b - \varepsilon$ za vsak $x \in A$ (z m smo označili Lebesguovo mero). Množica A je unija dveh množic, A_+ in A_-, pri čemer za vsak $x \in A_+$ velja $f'(x) \geq b - \varepsilon$ in za vsak $x \in A_-$ velja $f'(x) \leq -b + \varepsilon$. Vsaj ena od njiju ima pozitivno mero in brez škode za splošnost lahko privzamemo, da je to A_+. Ker je Lebesguova mera zunanje regularna, obstaja taka odprta množica $U \subset \mathbb{R}$, da je $A \subset U$ in:

$$m(U) < \frac{m(A_+)}{(1 - \varepsilon)}$$

Množica U je disjunktna unija največ števne unije odprtih intervalov U_λ. Trdimo: obstaja tak λ, da je $m(U_\lambda \cap A_+) > (1 - \varepsilon)m(U_\lambda)$. Če namreč ne bi bilo tako, bi veljalo:

$$m(A_+) = \sum_\lambda m(U_\lambda \cap A_+) \leq (1 - \varepsilon) \sum_\lambda m(U_\lambda) = (1 - \varepsilon)m(U)$$

kar ni res. Naj bo torej $m(U_\lambda \cap A_+) > (1 - \varepsilon)m(U_\lambda)$. Toda množica U_λ je odprt interval, torej $U_\lambda = (x, y)$. Sledi:

$$\begin{aligned} f(y) - f(x) = \int_x^y f'(t)\, dt = \int_{U_\lambda \cap A_+} f'(t)\, dt + \int_{U_\lambda \setminus A_+} f'(t)\, dt > \\ > (b - \varepsilon)(1 - \varepsilon)(y - x) - b\varepsilon(y - x) \end{aligned}$$

Dokazali smo:

$$a \geq \frac{f(x) - f(y)}{x - y} > (b - \varepsilon)(1 - \varepsilon) - b\varepsilon$$

Ko pošljemo ε proti 0, dobimo, da je tudi $a \geq b$. ■

Opomba. V resnici bomo pri nadaljnjem dokazovanju potrebovali le lažji del trditve, t. j. $a \leq b$.

Vrnimo se zdaj k naši funkciji g. Ker je $M_1 < \infty$, je g Lipschitzova funkcija, torej obstaja $\mathbf{E}(g(X))$. Še več, velja:

$$\mathbf{E}\big(g(X)\big) = \mathbf{E}\big(g(X')\big) = \mathbf{E}\big(g(X + D_K)\big)$$

Izraz na desni bomo funkcijo g razvili v Taylorjevo vrsto. Za ostanek bomo uporabili integralski obrazec. Velja torej:

$$\begin{aligned} g(x+h) &= g(x) + hg'(x) + \int_0^h (h-t)\, g''(x+t)\, dt = \\ &= g(x) + hg'(x) + \tfrac{1}{2}h^2 g''(x) + \int_0^h (h-t)(g''(x+t) - g''(x))\, dt \end{aligned}$$

Sledi:

$$\begin{aligned} \left|g(x+h) - g(x) - hg'(x) - \tfrac{1}{2}h^2 g''(x)\right| &\leq \int_0^h \left|(h-t)(g''(x+t) - g''(x))\right|\, dt \leq \\ &\leq M_3(g) \textstyle\int_0^h \left|t(h-t)\right|\, dt = \tfrac{1}{6}M_3(h)|h|^3 \end{aligned}$$

Če zdaj vstavimo $x := X$ in $h := D_K$, dobimo:

$$\begin{aligned} &\left|\mathbf{E}\left(D_K g'(X) + \tfrac{1}{2}D_K^2 g''(X)\right)\right| = \\ &= \left|\mathbf{E}\left(g(X + D_K) - g(X) - D_K g'(X) - \tfrac{1}{2}D_K^2 g''(X)\right)\right| \leq \tfrac{1}{6}M_3(g)\mathbf{E}\big(|D_K|^3\big) = \\ &= \tfrac{1}{6}n^{-3/2}M_3(g)\mathbf{E}\big(|Y_K - X_K|^3\big) \leq \tfrac{1}{6}n^{-3/2}M_3(g)\mathbf{E}\big((|Y_K| + |X_K|)^3\big) = \\ &= \tfrac{1}{6}n^{-3/2}M_3(g)\Big(\mathbf{E}\big(|Y_K|^3\big) + 3\mathbf{E}\big(|Y_K|^2\big)\mathbf{E}(|X_K|) + 3\mathbf{E}(|Y_K|)\mathbf{E}\big(|X_K|^2\big) + \mathbf{E}\big(|X_K|^3\big)\Big) = \\ &= n^{-3/2}M_3(g)\big(\alpha + \tfrac{1}{3}\beta\big) \leq \tfrac{4}{3}n^{-3/2}M_3(g)\beta \end{aligned}$$

Upoštevali smo, da so vse slučajne spremenljivke X_K in Y_K porazdeljene enako – tako kot X_1 in tudi, da sta slučajni spremenljivki X_K in Y_K neodvisni. Vse to dokažemo s podobno tehniko kot v verigi enakosti (2.1.1). Če torej označimo:

$$R_1 := \mathbf{E}\left(D_K g'(X) + \tfrac{1}{2}D_K^2 g''(X)\right)$$

velja ocena:

$$|R_1| \leq n^{-3/2}M_3(g)\big(\alpha + \tfrac{1}{3}\beta\big)$$

Izraz R_1 bomo zdaj izrazili z X, kar se morda ne bo dalo izraziti, pa bomo ocenili. Najprej velja:

$$\begin{aligned} \mathbf{E}(D_K g'(X)) &= \frac{1}{\sqrt{n}}\Big(\mathbf{E}(Y_K g'(X)) - \mathbf{E}(X_K g'(X))\Big) = \\ &= \frac{1}{\sqrt{n}}\left(\mathbf{E}(Y_K)\mathbf{E}(g'(X)) - \sum_{k=1}^n \mathbf{P}[K=k]\mathbf{E}(X_k g'(X))\right) = \\ &= -\frac{1}{n\sqrt{n}}\mathbf{E}\left(\sum_{k=1}^n X_k g'(X)\right) = -\frac{1}{n}\mathbf{E}(Xg'(X)) \end{aligned}$$

Nadalje je:

$$\begin{aligned} \mathbf{E}\left(\tfrac{1}{2}D_K^2 g''(X)\right) &= \frac{1}{2n}\Big(\mathbf{E}(Y_K^2)\mathbf{E}(g''(X)) - 2\mathbf{E}(Y_K)\mathbf{E}\big(X_K g''(X)\big) + \mathbf{E}\big(X_K^2 g''(X)\big)\Big) = \\ &= \frac{1}{2n}\Big(\mathbf{E}(g''(X)) + \mathbf{E}\big(X_K^2 g''(X)\big)\Big) \end{aligned}$$

V členu na skrajni desni slučajna spremenljivka X_K seveda ni neodvisna od X, je pa neodvisna od $X + D_K = X_1 + \ldots + X_{K-1} + Y_K + X_{K+1} + \ldots X_n$ (eksaktno to dokažemo z verigo enakosti, podobno verigi (2.1.1)). Zato je:

$$\mathbf{E}\big(X_K^2 g''(X + D_K)\big) = \mathbf{E}(X_K^2)\mathbf{E}(g''(X + D_K)) = \mathbf{E}(g''(X + D_K)) = \mathbf{E}(g''(X))$$

Če torej označimo:

$$R_2 := \frac{1}{2n}\mathbf{E}\big(X_K^2(g''(X + D_K) - g''(X))\big)$$

velja:

$$\mathbf{E}\left(\tfrac{1}{2}D_K^2 g''(X)\right) + R_2 = \frac{1}{2n}\Big(\mathbf{E}(g''(X)) + \mathbf{E}(g''(X))\Big) = \frac{1}{n}\mathbf{E}(g''(X))$$

Sledi:

$$R_1 + R_2 = \frac{1}{n}\mathbf{E}\big(g''(X) - Xg'(X)\big)$$

Ocenimo še izraz R_2. Iz $|g''(X + D_K) - g''(X)| \le M_3(g)|D_K|$ sledi:

$$\begin{aligned}|R_2| &\le \frac{1}{2n}M_3(g)\mathbf{E}\big(X_K^2|D_K|\big) \le \frac{1}{2n^{3/2}}M_3(g)\mathbf{E}\big(X_K^2(|Y_K| + |X_K|)\big) \le \\ &\le \frac{1}{2n^{3/2}}M_3(g)\Big(\mathbf{E}(X_K^2)\mathbf{E}(|Y_K|) + \mathbf{E}(|X_K|^3) \le \frac{1}{2n^{3/2}}M_3(g)(\alpha + \beta)\end{aligned}$$

Sledi:

$$|R_1 + R_2| \le n^{-3/2}M_3(g)\big(\tfrac{3}{2}\alpha + \tfrac{5}{6}\beta\big)$$

oziroma:

$$\big|\mathbf{E}\big(g''(X) - Xg'(X)\big)\big| \le n^{-1/2}M_3(g)\big(\tfrac{3}{2}\alpha + \tfrac{5}{6}\beta\big) \le \tfrac{7}{3}n^{-1/2}M_3(g)\beta \tag{3.1.2}$$

Opomba. V rokopisu [3] je zaradi majhnega spodrsljaja navedena nekoliko slabša ocena, kot je v (3.1.2): namesto konstante $\frac{7}{3}$ stoji $\frac{10}{3}$.

Izraz pod absolutno vrednostjo na levi je enak $\mathbf{E}\big((Gg)(X)\big)$, kjer je G generator Ornstein–Uhlenbeckovega procesa (glej izrek B.13.1), ki ima v resnici za ravnovesno porazdelitev normalno porazdelitev (glej stran 138). Rešiti je torej treba Steinovo enačbo:

$$Gg = f - \int f\,d\gamma_1 \tag{3.1.3}$$

kjer je $\gamma_1 = \mathrm{N}(0, 1)$ standardna normalna porazdelitev, pa bomo imeli oceno:

$$\left|\mathbf{E}\big(f(X)\big) - \int f\,d\gamma_1\right| \le n^{-1/2}M_3(g)\big(\tfrac{3}{2}\alpha + \tfrac{5}{6}\beta\big)$$

Steinovo enačbo bi lahko rešili tako kot v izreku 1.0.1. To bomo storili pri večrazsežni normalni aproksimaciji, tu pa bomo izrek 1.0.1 obšli in raje ravnali podobno kot pri Poissonovi aproksimaciji. Izraz Gg je namreč odvisen le od odvoda g'. Iz ocene (3.1.2) zdaj dobimo, da za vsako zvezno odvedljivo funkcijo h velja:

$$\big|\mathbf{E}\big(h'(X) - Xh(X)\big)\big| \le n^{-1/2}M_2(h)\big(\tfrac{3}{2}\alpha + \tfrac{5}{6}\beta\big) \le \tfrac{7}{3}n^{-1/2}M_2(h)\beta \tag{3.1.4}$$

Rešiti je torej treba prirejeno Steinovo enačbo:

$$h'(x) - xh(x) = f(x) - \int f \, d\gamma_1 \tag{3.1.5}$$

in oceniti $M_2(h)$, pa bomo imeli oceno za $\mathbf{E}\big(f(X)\big) - \int f \, d\gamma_1$. Če bomo znali globalno oceniti $M_2(h)$, ko f preteče določen razred testnih funkcij, bomo dobili oceno v primerni metriki (glej dodatek A). V naslednjem razdelku bomo izpeljali oceno v Wassersteinovi metriki. Le-ta nam bo potem dala tudi šibko konvergenco.

Bistvena prednost Chen–Steinove metode je, da deluje tudi za odvisne slučajne spremenljivke. Pri njih je glavni prijem *transformacija porazdelitve*: porazdelitvi μ priredimo porazdelitev μ^*. V članku [12] je opisana transformacija, ki vsaki porazdelitvi μ na $[0, \infty)$, za katero je $\lambda := \int x \, \mu(dx) < \infty$, priredi tako porazdelitev μ^* (*size-biased distribution*), da za vsako merljivo funkcijo h, za katero spodnja izraza obstajata, velja:

$$\int x \, h(x) \, \mu(dx) = \lambda \int h \, d\mu^*$$

Če ima torej W porazdelitev μ in W^* porazdelitev μ^*, velja:

$$\mathbf{E}\big(W h(W)\big) = \lambda \mathbf{E}\big(h(W^*)\big) \tag{3.1.6}$$

Če torej želimo aproksimirati porazdelitev dane slučajne spremenljivke X, najprej poiščemo slučajno spremenljivko X^*, ki ima tako transformirano porazdelitev, in sicer *na istem verjetnostnem prostoru*. Nato lahko zapišemo:

$$\mathbf{E}\big(h'(X) - (X - \lambda)h(X)\big) = \mathbf{E}(h'(X)) - \lambda \mathbf{E}\big(h(X^*) - h(X)\big)$$

Namesto faktorja X nastopa faktor $X - \lambda$, to pa zato, ker slučajna spremenljivka X ni centrirana. V drugem členu zdaj funkcijo h razvijemo v Taylorjevo vrsto. Če se X in X^* ne razlikujeta preveč, lahko vse skupaj dobro ocenimo.

V članku [12] je navedenih kar nekaj primerov uporabe. Tako se da podobno kot pri Poissonovi aproksimaciji tudi tu gledati lokalno odvisnost, ki naravno nastopa v slučajnih grafih.

Pomanjkljivost slednje metode pa je, da lepo deluje le za nenegativne slučajne spremenljivke ali kvečjemu za slučajne spremenljivke, ki so negativne le z zelo majhno verjetnostjo. Izrazito neneravna pa je uporaba te metode pri simetričnih porazdelitvah. Pri teh pa je primernejša druga metoda (*zero-bias transformation*), ki je opisana v članku [11]. Pri tej transformaciji namesto (3.1.6) velja:

$$\mathbf{E}\big(W h(W)\big) = \sigma^2 \mathbf{E}\big(h'(W^*)\big) \tag{3.1.7}$$

kjer ima slučajna spremenljivka W matematično upanje 0 in varianco σ. V članku [11] je naveden tudi primer uporabe v statistiki.

Na sploh je konstrukcija slučajne spremenljivke X^* na prvi ali drugi način močno odvisna od problema, ki ga rešujemo. Primer druge konstrukcije (zero-biased transformation) za neodvisne slučajne spremenljivke bomo navedli pri dokazu Berry–Esseenovega izreka.

3.2 Ocenjevanje v Wassersteinovi metriki

Recimo, da smo za dano slučajno spremenljivko X dobili oceno oblike:

$$\left|\mathbf{E}\big(h'(X) - Xh(X)\big)\right| \le c_0\, M_0(h) + c_1\, M_1(h) + c_2\, M_2(h)$$

kjer je:

$$M_0(h) := \sup_x |h(x)|$$

Če je npr. X normirana vsota neodvisnih in enako porazdeljenih slučajnih spremenljivk, iz formule (3.1.4) sledi, da lahko postavimo $c_0 = c_1 = 1$ in $c_2 = n^{-1/2}(\frac{5}{6}\alpha + 32\beta)$. Naš cilj je oceniti Wassersteinovo razdaljo med porazdelitvijo slučajne spremenljivke X in standardno normalno porazdelitvijo. Za vsako Lipschitzovo funkcijo f s konstanto 1 moramo torej rešiti Steinovo enačbo (3.1.5):

$$h'(x) - xh(x) = f(x) - \int f\, d\gamma_1 \tag{3.2.1}$$

in oceniti $M_0(h)$, $M_1(h)$ in $M_2(h)$, npr. z M_0, M_1 in M_2. Tako bomo dobili oceno:

$$d_{\mathrm{W}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) \le c_0 M_0 + c_1 M_1 + c_2 M_2$$

Lotimo se zdaj enačbe (3.2.1). To je nehomogena linearna diferencialna enačba prvega reda. Če označimo $\varphi := f - \int f\, \gamma_1$, ima ta enačba naslednjo splošno rešitev:

$$h(x) = e^{\frac{1}{2}x^2}\left(\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\varphi(t)\, dt + C\right) \tag{3.2.2}$$

Izmed konstant C je le ena taka, da je funkcija h omejena. Če sta namreč h_1 in h_2 rešitvi, ki propadata konstantama C_1 in C_2, velja $h_2 - h_1 = e^{\frac{1}{2}x^2}$. Ta funkcija pa seveda ni omejena in prav tako so neomejeni tudi vsi njeni odvodi. Dokazali pa bomo, da so pri $C = 0$ vse količine $M_0(h)$, $M_1(h)$ in $M_2(h)$ omejene. Ocenili jih bomo tako, da bomo funkcijo h izrazili z *odvodom* funkcije f, ki skoraj povsod obstaja, saj je f Lipschitzova. Še več, f je integral svojega odvoda.

Idejo ocene povzemamo po [24]. Naj bo torej:

$$h(x) = e^{\frac{1}{2}x^2}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\varphi(t)\, dt \tag{3.2.3}$$

Označimo:

$$\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt$$

Najprej velja:

$$\begin{aligned}
\varphi(x) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\big(f(x) - f(y)\big)e^{-\frac{1}{2}y^2}\, dy = \\
&= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{x}\int_{y}^{x} f'(z)\, e^{-\frac{1}{2}y^2} dz\, dy - \int_{x}^{\infty}\int_{x}^{y} f'(z)\, e^{-\frac{1}{2}y^2} dz\, dy\right) = \\
&= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{x} f'(z)\int_{-\infty}^{z} e^{-\frac{1}{2}y^2} dy\, dz - \int_{x}^{\infty} f'(z)\int_{z}^{\infty} e^{-\frac{1}{2}y^2} dy\, dz\right) = \\
&= \int_{-\infty}^{x} f'(z)\, \Phi(z)\, dz - \int_{x}^{\infty} f'(z)\big(1 - \Phi(z)\big)\, dz
\end{aligned} \tag{3.2.4}$$

Sledi:

$$
\begin{aligned}
h(x) &= e^{\frac{1}{2}x^2}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\varphi(t)\,dt = \\
&= e^{\frac{1}{2}x^2}\int_{-\infty}^{x}\left(\int_{-\infty}^{t} f'(z)\Phi(z)\,dz - \int_{t}^{\infty} f'(z)\big(1-\Phi(z)\big)\,dz\right)e^{-\frac{1}{2}t^2}dt = \\
&= e^{\frac{1}{2}x^2}\left(\int_{-\infty}^{x} f'(z)\Phi(z)\int_{z}^{x} e^{-\frac{1}{2}t^2}dt\,dz - \int_{-\infty}^{x} f'(z)\big(1-\Phi(z)\big)\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}dt\,dz - \right.\\
&\quad\left. - \int_{x}^{\infty} f'(z)\big(1-\Phi(z)\big)\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}dt\,dz\right) = \\
&= \sqrt{2\pi}\,e^{\frac{1}{2}x^2}\left(\int_{-\infty}^{x} f'(z)\Phi(z)\big(\Phi(x)-\Phi(z)\big)dz - \int_{-\infty}^{x} f'(z)\big(1-\Phi(z)\big)\Phi(z)\,dz - \right.\\
&\quad\left. - \int_{x}^{\infty} f'(z)\big(1-\Phi(z)\big)\Phi(x)\,dz\right) = \\
&= -\sqrt{2\pi}\,e^{\frac{1}{2}x^2}\left(\big(1-\Phi(x)\big)\int_{-\infty}^{x} f'(z)\,\Phi(z)\,dz + \Phi(x)\int_{x}^{\infty} f'(z)\big(1-\Phi(z)\big)dz\right)
\end{aligned}
\tag{3.2.5}
$$

Ker je f Lipschitzova s konstanto 1, je $|f'(z)| \le 1$. Sledi:

$$
|h(x)| \le \sqrt{2\pi}\,e^{\frac{1}{2}x^2}\left(\big(1-\Phi(x)\big)\int_{-\infty}^{x}\Phi(z)\,dz + \Phi(x)\int_{x}^{\infty}\big(1-\Phi(z)\big)dz\right)
$$

Izraz na desni pa ni nič drugega kot funkcija $h(x)$, če vstavimo $f(x) = -x$. V tem primeru pa po formuli (3.2.3) dobimo $h(x) = 1$. Za vsako Lipschitzovo funkcijo f s konstanto 1 je torej:

$$
M_0(h) = \sup_{x\in\mathbb{R}}|h(x)| \le 1
$$

in ta ocena je najboljša možna.

Funkcija h je seveda klasična rešitev enačbe (3.1.5), torej je zvezno odvedljiva in velja:

$$
h'(x) = \varphi(x) + xh(x) \tag{3.2.6}
$$

Med drugim to tudi pomeni, da so izpolnjeni pogoji trditve 3.1.1, torej velja $M_1(h) = \sup_x|h'(x)|$. Iz formul (3.2.4), (3.2.5) in (3.2.6) dobimo:

$$
\begin{aligned}
h'(x) = &\left(1-\sqrt{2\pi}\,x\,e^{\frac{1}{2}x^2}\big(1-\Phi(x)\big)\right)\int_{-\infty}^{x} f'(z)\,\Phi(z)\,dz - \\
&-\left(1+\sqrt{2\pi}\,x\,e^{\frac{1}{2}x^2}\Phi(x)\right)\int_{x}^{\infty} f'(z)\big(1-\Phi(z)\big)dz
\end{aligned}
\tag{3.2.7}
$$

Nedoločeni integral funkcije Φ se da eksplicitno izračunati. Z integracijo per partes dobimo:

$$
\int\Phi(x)\,dx = x\,\Phi(x) - \frac{1}{\sqrt{2\pi}}\int x\,e^{-\frac{1}{2}x^2}dx = x\Phi(x) + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} + C
$$

Sledi:

$$
\int_{-\infty}^{x}\Phi(z)\,dz = x\Phi(x) + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx \tag{3.2.8}
$$

Seveda se da integrirati tudi $1 - \Phi(x) = \Phi(-x)$. Iz zgornje formule izpeljemo:

$$\begin{aligned}\int_x^\infty \bigl(1 - \Phi(z)\bigr)\,dz = \int_x^\infty \Phi(-z)\,dz = -\int_{-x}^{-\infty} \Phi(z)\,dz = \int_{-\infty}^{-x} \Phi(z)\,dz = \\ = -\,x\Phi(-x) + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} = -x\bigl(1 - \Phi(x)\bigr) + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\end{aligned} \tag{3.2.9}$$

Iz formule (3.2.7) zdaj dobimo:

$$\begin{aligned}h'(x) = \sqrt{2\pi}\, e^{\frac{1}{2}x^2} \biggl(\int_x^\infty \bigl(1 - \Phi(z)\bigr)\,dz \cdot \int_{-\infty}^x f'(z)\,\Phi(z)\,dz - \\ - \int_{-\infty}^x \Phi(z)\,dz \cdot \int_x^\infty f'(z)\bigl(1 - \Phi(z)\bigr)\,dz \biggr)\end{aligned}$$

Sledi:

$$|h'(x)| \le 2\sqrt{2\pi}\, e^{\frac{1}{2}x^2} \int_{-\infty}^x \Phi(z)\,dz \cdot \int_x^\infty \bigl(1 - \Phi(z)\bigr)\,dz =: F(x) \tag{3.2.10}$$

Zgornja ocena je dosežena pri $f(z) = |z - x|$, torej je maksimum funkcije F najboljša možna ocena za $\sup_h M_1(h)$, kjer je h definirana po formuli (3.2.3), ko f preteče vse Lipschitzove funkcije s konstanto 1. Velja:

$$\begin{aligned}F(-x) = 2\sqrt{2\pi}\, e^{\frac{1}{2}x^2} \int_{-\infty}^{-x} \Phi(z)\,dz \cdot \int_{-x}^\infty \bigl(1 - \Phi(z)\bigr)\,dz = \\ = 2\sqrt{2\pi}\, e^{\frac{1}{2}x^2} \int_x^\infty \Phi(-z)\,dz \cdot \int_{-\infty}^x \bigl(1 - \Phi(-z)\bigr)\,dz = F(x)\end{aligned}$$

torej je funkcija F soda, zato je $F(x)$ dovolj oceniti za $x \ge 0$. Za tak x pa velja:

$$\int_{-\infty}^x \Phi(z)\,dz = \tfrac{1}{2} + \textstyle\int_0^x \Phi(z)\,dz \le \tfrac{1}{2} + x$$

$$\begin{aligned}\sqrt{2\pi}\, e^{\frac{1}{2}x^2} \int_x^\infty \bigl(1 - \Phi(z)\bigr)\,dz = e^{\frac{1}{2}x^2} \int_x^\infty \int_z^\infty e^{-\frac{1}{2}t^2}\,dt\,dz = \int_x^\infty e^{-\frac{1}{2}(t^2 - x^2)} \int_x^t dz\,dt = \\ = \int_x^\infty (t - x)\, e^{-\frac{1}{2}(t-x)(t+x)}\,dt = \int_0^\infty s\, e^{-\frac{1}{2}s^2 - sx}\,ds\end{aligned}$$

Za vsak $x \ge 0$ torej velja:

$$F(x) \le (1 + 2x) \int_0^\infty s\, e^{-\frac{1}{2}s^2 - sx}\,ds$$

Slednji integral pa lahko ocenimo na dva načina. Po eni strani velja:

$$\int_0^\infty s\, e^{-\frac{1}{2}s^2 - sx}\,ds \le \int_0^\infty s\, e^{-\frac{1}{2}s^2}\,ds = 1$$

Za $0 \le x \le 1$ torej velja $F(x) \le 1 + 2x \le 3$. Po drugi strani pa velja tudi:

$$\int_0^\infty s\, e^{-\frac{1}{2}s^2 - sx}\,ds \le \int_0^\infty s\, e^{-sx}\,ds = \frac{1}{x^2}$$

Za $x \geq 1$ potem velja $F(x) \leq 1 + 2x\,x^{-2} \leq 3$. Dokazali smo, da za vsak $x \geq 0$ in zaradi sodosti potem tudi za vsak realen x velja $F(x) \leq 3$. Brž ko je torej f Lipschitzova funkcija s konstanto 1 in je funkcija h definirana po formuli (3.2.3), potem velja:

$$M_1(h) \leq 3$$

Ocenimo zdaj še $M_2(h)$. Iz formule (3.2.6) sledi, da je h' na vsakem omejenem intervalu Lipschitzova, torej skoraj povsod odvedljiva in integral svojega odvoda. Natančneje, h' je odvedljiva natanko tam, kjer je odvedljiva φ, in tam velja:

$$h''(x) = \varphi'(x) + h(x) + x\,h'(x) = \varphi'(x) + x\,\varphi(x) + (1 + x^2)h(x)$$

Iz formul (3.2.4) in (3.2.5) sledi:

$$\begin{aligned} h''(x) = \varphi'(x) &+ \Big(x - \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\big(1 - \Phi(x)\big)\Big) \int_{-\infty}^{x} f'(z)\,\Phi(z)\,dz - \\ &- \Big(x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\Phi(x)\Big) \int_{x}^{\infty} f'(z)\big(1 - \Phi(z)\big)dz \end{aligned} \tag{3.2.11}$$

Lema 3.2.1. *Za vsak $x \in \mathbb{R}$ velja:*

$$\begin{aligned} -\,x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\big(1 - \Phi(x)\big) &\geq 0 \\ x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\Phi(x) &\geq 0 \end{aligned}$$

Dokaz. Dovolj je dokazati le prvo neenačbo, saj drugo dobimo iz nje tako, da x zamenjamo z $-x$. Prva neenačba očitno drži za $x \leq 0$, torej smemo privzeti, da je $x > 0$. V tem primeru pa je prva neenačba ekvivalentna neenačbi:

$$\frac{x}{1+x^2}\,e^{-\frac{1}{2}x^2} \leq \int_x^\infty e^{-\frac{1}{2}t^2}dt$$

Odvajajmo izraz na levi. Po krajšem računu dobimo:

$$\left(\frac{x}{1+x^2}\,e^{-\frac{1}{2}x^2}\right)' = -\frac{(1+x^2)^2 - 2}{(1+x^2)^2}\,e^{-\frac{1}{2}x^2}$$

Če zdaj neenačbo:

$$\frac{(1+t^2)^2 - 2}{(1+t^2)^2}e^{-\frac{1}{2}t^2} \leq e^{-\frac{1}{2}t^2}$$

integriramo od x do neskončno, dobimo natanko to, kar smo iskali. ■

Iz dejstva, da je f Lipschitzova s konstanto 1, prejšnje leme ter formul (3.2.8) in (3.2.9) zdaj iz formule (3.2.11) sledi:

$$
\begin{aligned}
h''(x) \leq 1 &+ \left(-x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\big(1-\Phi(x)\big)\right)\int_{-\infty}^{x}\Phi(z)\,dz + \\
&+ \left(x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\Phi(x)\right)\int_{x}^{\infty}\big(1-\Phi(z)\big)dz = \\
= 1 &+ \left(-x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\big(1-\Phi(x)\big)\right)\left(x\,\Phi(x) + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right) + \\
&+ \left(x + \sqrt{2\pi}(1+x^2)e^{\frac{1}{2}x^2}\Phi(x)\right)\left(-x\big(1-\Phi(x)\big) + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right) = \\
= 2 &
\end{aligned}
$$

Dokazali smo torej oceno:

$$M_2(h) = \operatorname*{ess\,sup}_{x\in\mathbb{R}}|h''(x)| \leq 2$$

Seveda tudi tu nastane vprašanje, ali je ta ocena najboljša možna. V formuli (3.2.11) je maksimum absolutne vrednosti izraza na desni brez $\varphi'(x)$ dosežen, če je odvod $\varphi' = f'$ konstanta 1 ali -1. Toda v tem primeru se ta del izraza uniči s $\varphi'(x)$. Če je torej $\varphi' = -1$, je drugi del izraza enak 1, kar se uniči s $\varphi'(x) = -1$. Toda funkcijo φ' lahko tudi malo popravimo. Izberimo $a \in \mathbb{R}$ in $n \in \mathbb{N}$. Obstaja taka funkcija φ_n, da je $\int \varphi_n \, d\gamma_1 = 0$ in da velja:

$$\varphi_n'(x) = \begin{cases} 1 & |x-a| \leq \frac{1}{n} \\ -1 & \text{sicer} \end{cases}$$

Naj bo $f_n = \varphi_n$ in naj bodo h_n funkcije, izvedene iz f_n po formuli (3.2.3). Tedaj iz formule (3.2.11) in izreka o dominirani konvergenci sledi:

$$\lim_{n\to\infty} h_n''(a) = 2$$

Za vsak $\varepsilon > 0$ torej obstaja tak $n \in \mathbb{N}$, da je $h_n''(a) > 2 - \varepsilon/2$. Toda drugi del izraza na desni v formuli (3.2.11) je zvezna funkcija, prvi del (namreč $\varphi_n''(x)$) pa je na intervalu $(a - \frac{1}{n}, a + \frac{1}{n})$ konstanten. Zato obstaja tudi tak $\delta > 0$, da za vsak x, za katerega je $|x - a| < \delta$, velja $|h_n''(x)| > 2 - \varepsilon$. To pa pomeni, da je $M_2(h) = \operatorname{ess\,sup}_x |h_n''(x)| > 2 - \varepsilon$. Konstanta 2 je torej v resnici najboljša zgornja meja.

Zgornje rezultate lahko povzamemo v naslednjo trditev.

Trditev 3.2.2. *Če je f Lipschitzova funkcija s konstanto 1 in je h rešitev Steinove enačbe (3.1.5), definirana po formuli (3.2.3) (kar je edina omejena rešitev te enačbe), je h odvedljiva, odvod h' je absolutno zvezen in integral svojega odvoda, veljajo pa tudi ocene:*

$$M_0(h) \leq 1, \quad M_1(h) \leq \sup_{x\in\mathbb{R}} F(x) \leq 3, \quad M_2(h) \leq 2$$

kjer je funkcija F definirana po formuli (3.2.10). Vse te ocene, razen ocene $\sup_{x\in\mathbb{R}} F(x) \leq 3$, so tudi najboljše možne.

Opomba. Zgornje ocene so omenjene tudi v [3]. Tam je napisano, da ni težko videti, da je $M_0(h) \leq 2$, $M_1(h) \leq 3$ in $M_2(h) \leq 6$. Naši oceni za $M_0(h)$ in $M_2(h)$ sta torej boljši (celo najboljši možni), pa tudi ocena za $M_1(h)$ bi se dala še izboljšati.

Iz naših ocen sledi naslednji osnovni izrek enorazrežne normalne aproksimacije po Chen–Steinovi metodi v Wassersteinovi metriki.

Izrek 3.2.3. *Naj bo X slučajna spremenljivka in naj za vsako odvedljivo funkcijo $h\colon \mathbb{R} \to \mathbb{R}$, katere odvod je Lipschitzova funkcija, velja:*

$$\left|\mathbf{E}\big(h'(X) - Xh(X)\big)\right| \le c_0 M_0(h) + c_1 M_1(h) + c_2 M_2(h)$$

Tedaj velja:

$$d_{\mathrm{W}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) \le c_0 + 3c_1 + 2c_2$$

Posledica 3.2.4. *Naj bodo $X_1, \dots X_n$ neodvisne slučajne spremenljivke. Naj bo $\mathbf{E}(X_k) = 0$, $\mathrm{var}(X_k) = 1$, $\mathbf{E}(|X_k|) = \alpha$ in $\mathbf{E}(|X_k|^3) = \beta$. Naj bo $X := n^{-1/2} \sum_{k=1}^n X_k$. Tedaj velja:*

$$d_{\mathrm{W}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) \le \frac{3\alpha + \frac{5}{3}\beta}{\sqrt{n}} \le \frac{14}{3} \frac{\beta}{\sqrt{n}}$$

Opomba. Iz zgornje posledice sledi tudi centralni limitni izrek, a seveda le za primer, ko imajo slučajne spremenljivke končen tretji moment. Bistvena prednost posledice 3.2.4 pa je seveda ocena napake.

3.3 Berry–Esseenov izrek

V prejšnjem razdelku smo porazdelitev normalizirane vsote neodvisnih in enako porazdeljenih slučajnih spremenljivk aproksimirali standardno normalno porazdelitvijo. Napako smo ocenili v Wassersteinovi metriki. V tem razdelku pa bomo dokazali Berry–Esseenov izrek, ki nam napako oceni v metriki Kolmogorova (glej dodatek A, stran 75).

Izrek 3.3.1 (Berry, Esseen). *Naj bodo $X_1, \dots X_n$ neodvisne in enako porazdeljene slučajne spremenljivke. Naj bo $\mathbf{E}(X_k) = 0$ in $\mathrm{var}(X_k) = 1$. Naj bo tudi $\beta := \mathbf{E}(|X_k|^3) < \infty$. Označimo $X := n^{-1/2} \sum_{k=1}^n X_k$. Tedaj velja:*

$$d_{\mathrm{K}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) = \sup_x \left|\mathbf{P}[X \le x] - \Phi(x)\right| \le \frac{C\beta}{\sqrt{n}}$$

Tu je $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2}\, dt$, C pa je univerzalna konstanta, za katero je znano, da je $\frac{1}{2\pi} \le C \le 0.8$.

Pri dokazovanju Berry–Esseenovega izreka se žal ne moremo sklicevati na rezultate prejšnjega razdelka, saj testne funkcije:

$$f_z := I_{(-\infty,z]}, \text{ t. j. } f_z(x) = I[x \le z] = \begin{cases} 1 & x \le z \\ 0 & x > z \end{cases}$$

niso Lipschitzove. Še huje pa je, da se ne moremo sklicati niti na rezultate prvega razdelka, t. j. oceno (3.1.4), saj ima odvod rešitve Steinove enačbe (3.1.5), t. j.:

$$h'(x) - xh(x) = f_z - \int f_z \, d\gamma_1 = I[x \le z] - \Phi(z) \tag{3.3.1}$$

kjer je:

$$\Phi(z) = \gamma_1((-\infty, z]) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$$

skok pri $x = z$. Tudi samo Steinovo enačbo moramo obravnavati nekoliko bolj pazljivo. Prej jo je bilo dovolj gledati na klasičen način, t. j. iskati vse zvezno odvedljive funkcije, ki rešijo enačbo. Tu pa moramo gledati vse funkcije, ki so absolutno zvezne in integral svojega odvoda ter rešijo zgornjo enačbo za skoraj vsak x. Gledano na ta način, funkcija h_z, podana s formulo (3.2.3):

$$h_z(x) := e^{\frac{1}{2}x^2} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} \varphi_z(t)\, dt = -e^{\frac{1}{2}x^2} \int_{x}^{\infty} e^{-\frac{1}{2}t^2} \varphi_z(t)\, dt \tag{3.3.2}$$

kjer je $\varphi_z = f_z - \Phi(z)$, reši Steinovo enačbo (3.3.1). Povsod razen v z je h_z neskončno gladka in za vsak $x \neq z$ velja (3.3.1). V točki z pa ima h' skok, ki je enak:

$$\begin{aligned} \lim_{x \downarrow z} h_z'(x) - \lim_{x \uparrow z} h_z'(x) &= \lim_{x \downarrow z} \big(x h_z(x) + \varphi_z(x)\big) - \lim_{x \uparrow z} \big(x h_z(x) + \varphi_z(x)\big) = \\ &= z h_z(z) - \Phi(z) - \big(z h_z(z) + 1 - \Phi(z)\big) = -1 \end{aligned} \tag{3.3.3}$$

Torej je $M_2(h) = \infty$, zato res ne moremo uporabiti rezultatov prvega razdelka. Toda zunaj skoka drugi odvod obstaja in ga je mogoče omejiti, prav tako pa je mogoče omejiti tudi prvi odvod in seveda samo funkcijo h_z.

Lema 3.3.2. *Velja:*

(1) $0 \leq h_z \leq \frac{\sqrt{2\pi}}{4}$.

(2) *Za* $x < z$ *velja* $0 \leq h_z'(x) \leq 1$, *za* $x > z$ *pa* $-1 \leq h_z'(x) \leq 0$.

(3) *Za* $x \neq z$ *velja* $|h_z''(x)| \leq 1 + |x|$.

Dokaz. *(1)*: Za $x \leq z$ velja:

$$h_z(x) = \big(1 - \Phi(z)\big) e^{\frac{1}{2}x^2} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt = \sqrt{2\pi}\big(1 - \Phi(z)\big) e^{\frac{1}{2}x^2} \Phi(x) \tag{3.3.4}$$

za $x \geq z$ pa velja:

$$h_z(x) = \Phi(z)\, e^{\frac{1}{2}x^2} \int_{x}^{\infty} e^{\frac{1}{2}t^2}\, dt = \sqrt{2\pi} \Phi(z)\, e^{\frac{1}{2}x^2} \big(1 - \Phi(x)\big) \tag{3.3.5}$$

Očitno je torej $h_z(x) \geq 0$ za vsak $x \in \mathbb{R}$. Ker je funkcija Φ naraščajoča, je tudi:

$$h_z(x) \leq \sqrt{2\pi} e^{\frac{1}{2}x^2} \Phi(x) \big(1 - \Phi(x)\big) =: F(x)$$

Očitno je $F(0) = \frac{\sqrt{2\pi}}{4}$. Pokazati je torej treba, da funkcija F doseže svoj maksimum v točki 0. To bomo storili tako, da bomo F razvili v Taylorjevo vrsto. Le-ta očitno konvergira za vsak x, saj je F produkt celih analitičnih funkcij. Dokaz povzemamo iz [24].

Najprej bomo z indukcijo pokazali, da za vsak $k \geq 1$ velja:

$$F^{(k)}(x) = xF^{(k-1)}(x) + (k-1)F^{(k-2)}(x) + \left(\frac{d}{dx}\right)^{k-1} \big(1 - 2\Phi(x)\big) \tag{3.3.6}$$

Pri $k = 1$ v formuli nastopa $F^{(-1)}(x)$, ki naj bo npr. kaka primitivna funkcija funkcije F. Le-ta tako ali tako ne igra nobene vloge, saj je pomnožena z 0.

Najprej velja:

$$F'(x) = \sqrt{2\pi} x e^{\frac{1}{2}x^2} \Phi(x)\big(1 - \Phi(x)\big) + 1 - \Phi(x) - \Phi(x) = xF(x) + 1 - 2\Phi(x) \tag{3.3.7}$$

Za $k = 1$ torej formula (3.3.6) velja. Naredimo zdaj indukcijski korak s k na $k + 1$ tako, da formulo (3.3.6) odvajamo. Dobimo:

$$F^{(k+1)}(x) = xF^{(k-2)}(x) + kF^{(k-2)}(x) + \left(\frac{d}{dx}\right)^{k} \big(1 - 2\Phi(x)\big)$$

kar je ravno formula (3.3.6), pri čemer je k povečan za 1. Naj bo zdaj $F(x) = \sum_{k=0}^{\infty} a_k x^k$ razvoj funkcije F v Taylorjevo vrsto. Iz formule (3.3.6) zdaj sledi:

$$k!\, a_k = (k-1)! a_{k-2} + \left(\frac{d}{dx}\right)^{k-1} \bigg|_{x=0} \big(1 - 2\Phi(x)\big)$$

oziroma:

$$ka_k = a_{k-2} + \frac{1}{(k-1)!} \left(\frac{d}{dx}\right)^{k-1} \bigg|_{x=0} \big(1 - 2\Phi(x)\big) \tag{3.3.8}$$

Če integriramo eksponentno vrsto, dobimo:

$$1 - 2\Phi(x) = \sqrt{\frac{2}{\pi}} \sum_{r=1}^{\infty} (-1)^r \frac{x^{2r-1}}{2^{r-1}(r-1)!\,(2r-1)}$$

Če je k lih, iz formule (3.3.8) sledi $ka_k = a_{k-2}$. Ker po formuli (3.3.7) velja $a_1 F'(0) = 0$, je $a_k = 0$ za vse lihe k. Za sode indekse pa označimo $b_r := a_{2r}$. Dobimo:

$$2rb_r = b_{r-1} + \sqrt{\frac{2}{\pi}} \frac{(-1)^r}{2^{r-1}(r-1)!\,(2r-1)}$$

Po substituciji $c_r := 2^r r!\, b_r$ in množenju z $2^{r-1}(r-1)!$ dobimo:

$$c_r = c_{r-1} + \sqrt{\frac{2}{\pi}} \frac{(-1)^r}{2r-1}$$

Velja $c_0 = b_0 = a_0 = \frac{\sqrt{2\pi}}{4}$. Sledi:

$$c_r = \frac{\sqrt{2\pi}}{4} + \sum_{i=1}^{r} \sqrt{\frac{2}{\pi}} \frac{(-1)^i}{2i-1} = \sqrt{\frac{2}{\pi}} \left(\frac{\pi}{4} + \sum_{i=1}^{r} \frac{(-1)^i}{2i-1}\right)$$

Zdaj, ko smo izračunali koeficiente c_r, lahko zapišemo razvoj funkcije F v Taylorjevo vrsto. Velja:

$$F(x) = \sum_{r=0}^{\infty} \frac{c_r}{r!} \left(\frac{x^2}{2}\right)^r = F(0) + \sum_{r=1}^{\infty} \frac{c_r}{r!} \left(\frac{x^2}{2}\right)^r \tag{3.3.9}$$

Dokazati moramo, da je (vsaj za majhne x) preostanek vrste negativen. To bomo storili tako, da bomo koeficiente pobrali po dva in dva in pokazali, da je vsota vsakega para koeficientov negativna. Najprej opazimo, da so koeficienti c_r delne vsote alternirajoče vrste. Sledi:

$$c_1 \leq c_3 \leq \ldots \leq \lim_{r\to\infty} c_r \leq \ldots \leq c_2 \leq c_0$$

Ker je $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots = \frac{\pi}{4}$, je $\lim_{r\to\infty} c_r = 0$. Če je r sod, je torej $c_r > 0$, drugače pa je $c_r < 0$. Nadalje lahko zapišemo:

$$c_r = -\sqrt{\frac{2}{\pi}} \sum_{i=r+1}^{\infty} \frac{(-1)^i}{2i-1}$$

Naj bo r lih. Tedaj velja:

$$\begin{aligned}
\sqrt{\frac{\pi}{2}}(c_r + c_{r+1}) &= -\sum_{i=r+1}^{\infty} \frac{(-1)^i}{2i-1} - \sum_{i=r+2}^{\infty} \frac{(-1)^i}{2i-1} = \\
&= \sum_{i=r+1}^{\infty} (-1)^i \left(\frac{1}{2i+1} - \frac{1}{2i-1}\right) = -2\sum_{i=r+1}^{\infty} \frac{(-1)^i}{4i^2-1} = \\
&= -2\sum_{k=0}^{\infty} \left(\frac{1}{4(r+2k+1)^2-1} - \frac{1}{4(r+2k+2)^2-1}\right) < 0
\end{aligned}$$

Torej je $c_{2k} < -c_{2k-1}$ za vsak $k \in \mathbb{N}$. Iz formule (3.3.9) zdaj sledi:

$$\begin{aligned}
F(x) &= F(0) + \sum_{k=1}^{\infty} \left(\frac{c_{2k-1}}{(2k-1)!}\left(\frac{x^2}{2}\right)^{2k-1} + \frac{c_{2k}}{(2k)!}\left(\frac{x^2}{2}\right)^{2k}\right) \leq \\
&\leq F(0) + \sum_{k=1}^{\infty} \frac{c_{2k-1}}{(2k-1)!}\left(\frac{x^2}{2}\right)^{2k-1}\left(1 - \frac{x^2}{4k}\right)
\end{aligned}$$

Če je $|x| \leq 2$, je zdaj vsota na desni res negativna. Za $|x| \leq 2$ torej velja $F(x) \leq F(0) = \frac{\sqrt{2\pi}}{4}$. Ocenimo zdaj $F(x)$ še za $|x| \geq 2$. Najprej za vsak $x > 0$ velja:

$$\begin{aligned}
\sqrt{2\pi}\, e^{\frac{1}{2}x^2}\big(1 - \Phi(x)\big) &= \int_x^{\infty} e^{-\frac{1}{2}(t^2-x^2)}\, dt = \int_0^{\infty} e^{-\frac{1}{2}s^2 - sx}\, ds \leq \\
&\leq \int_0^{\infty} e^{-sx}\, ds = \frac{1}{x}
\end{aligned} \tag{3.3.10}$$

Sledi:

$$F(x) \leq \sqrt{2\pi}\, e^{\frac{1}{2}x^2}\big(1 - \Phi(x)\big) \leq \frac{1}{x}$$

Ker je funkcija F soda, potem za vsak $x \neq 0$ velja:

$$F(x) \leq \frac{1}{|x|} \tag{3.3.11}$$

Če je $|x| \geq 2$, je potemtakem $F(x) \leq \frac{1}{2} < \frac{\sqrt{2\pi}}{4}$. Torej je res:

$$F(x) \leq \frac{\sqrt{2\pi}}{4} \tag{3.3.12}$$

za vsak $x \in \mathbb{R}$, od tod pa že sledi točka (1).

(2): Iz formul (3.3.4) in (3.3.5) sledi, da je $h_{-z}(x) = h_z(-x)$, torej tudi $h'_{-z}(x) = -h'_z(-x)$. Od tod pa po kratkem premisleku sledi, da je trditev dovolj dokazati le za $z \geq 0$. Iz formule (3.3.4) sledi, da za $x < z$ velja:

$$h'_z(x) = \big(1 - \Phi(z)\big)\Big(1 + \sqrt{2\pi}\, x\, e^{\frac{1}{2}x^2}\Phi(x)\Big) \tag{3.3.13}$$

Naj bo $x < 0$. Če v oceni (3.3.10) x zamenjamo z $-x$, dobimo:

$$\sqrt{2\pi}\, e^{\frac{1}{2}x^2}\Phi(x) < -\frac{1}{x} \tag{3.3.14}$$

Sledi $0 \leq h'_z(x) \leq 1 - \Phi(z) \leq 1$. Naj bo zdaj $0 \leq x < z$. Tedaj je očitno $h'_z(x) \geq 0$. Iz dejstva, da je Φ naraščajoča funkcija, in ocene (3.3.10) pa sledi še:

$$h'_z(x) \leq 1 - \Phi(z) + \sqrt{2\pi}\, x\, e^{\frac{1}{2}x^2}\Phi(z)\big(1 - \Phi(x)\big) \leq 1 - \Phi(z) + \Phi(z) \leq 1$$

Za $x > z$ pa iz formule (3.3.5) dobimo:

$$h'_z(x) = \Phi(z)\Big(-1 + \sqrt{2\pi}\, x\, e^{\frac{1}{2}x^2}\big(1 - \Phi(x)\big)\Big) \tag{3.3.15}$$

Očitno je $h'_z(x) \geq -1$, iz ocene (3.3.10) pa sledi, da je tudi $h'_z(x) \leq 0$. To pa je vse, kar je bilo treba dokazati.

(3): Ker je $h_{-z}(x) = h_z(-x)$, je vse skupaj dovolj dokazati za primer, ko je $x < z$. V tem primeru iz formule (3.3.13) sledi:

$$h''_z(x) = \big(1 - \Phi(z)\big)\Big(\sqrt{2\pi}(1 + x^2)e^{\frac{1}{2}x^2}\Phi(x) + x\Big)$$

Ker je funkcija Φ naraščajoča, sledi:

$$|h''(z)| \leq \big(1 - \Phi(x)\big)\left|\sqrt{2\pi}(1 + x^2)e^{\frac{1}{2}x^2}\Phi(x) + x\right| = \left|(1 + x^2)F(x) + x\big(1 - \Phi(x)\big)\right|$$

Iz ocen (3.3.11) in (3.3.12) sledi $(1 + x^2)F(x) \leq \frac{\sqrt{2\pi}}{4} + |x|$. Naj bo $x \geq 0$. V tem primeru bomo za drugi člen pokazali, da je enak največ 0.373. To je očitno res za $0 \leq x \leq 0.373$. Za $x > 0.373$ pa po oceni (3.3.10) velja:

$$x\big(1 - \Phi(x)\big) \leq \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} < \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}0.373^2} < 0.373$$

Za $x \geq 0$ torej velja:

$$|h''z(x)| \leq \frac{\sqrt{2\pi}}{4} + x + 0.373 < 1 + x$$

Za $x \leq 0$ pa sta izraza $(1+x^2)F(x)$ in $x\big(1-\Phi(x)\big)$ nasprotnega predznaka. Ker sta oba po absolutni vrednosti enaka največ $\frac{\sqrt{2\pi}}{4} + |x| < 1 + |x|$, to velja tudi za njuno vsoto. Torej tudi za $x \leq 0$ velja $|h''_z(x)| \leq 1 + |x|$. ■

Lotimo se zdaj dokazovanja Berry–Esseenovega izreka. Podali bomo Steinov dokaz iz leta 1970, ki je objavljen v članku [14]. Naj bodo torej $X_1, \dots X_n$ kot v izreku 3.3.1. Označimo $Y_k := n^{-1/2} X_k$, tako da je $X = \sum_{k=1}^n Y_k$. Tudi slučajne spremenljivke Y_k so neodvisne in enako porazdeljene. Označimo še $X^* := \sum_{k=1}^{n-1} Y_k$. Naj bo h: $\mathbb{R} \to \mathbb{R}$ absolutno zvezna funkcija z omejenim odvodom (pravkar dokazana lema nam pove, da so take tudi funkcije h_z). Tedaj velja:

$$\begin{aligned}\mathbf{E}\big(Xh(X)\big) &= \sum_{k=1}^n \mathbf{E}\big(Y_k h(\textstyle\sum_{r\neq k} Y_r + Y_k)\big) = \\ &= n\mathbf{E}\big(Y_n h(X^* + Y_n)\big) = n\mathbf{E}\Big(Y_n\big(h(X^*+Y_n) - h(X^*)\big)\Big)\end{aligned}$$

Pri drugem enačaju smo upoštevali simetrijo, pri tretjem pa neodvisnost in dejstvo, da je $\mathbf{E}(Y_n) = 0$. V naslednjem koraku bomo upoštevali neodvisnost in matematično upanje zapisali kot dvojni integral. Naj bo μ porazdelitev slučajne spremenljivke X^*, ν pa porazdelitev slučajne spremenljivke Y_n. Tedaj velja:

$$\begin{aligned}\mathbf{E}\big(Xh(X)\big) &= n\int\int y\big(h(x+y) - h(x)\big)\,\nu(dy)\,\mu(dx) = \\ &= -n\int\int_{y\leq 0}\int_y^0 y\,h'(x+t)\,dt\,\nu(dy)\,\mu(dx) + \\ &\quad + n\int\int_{y>0}\int_0^y y\,h'(x+t)\,dy\,\nu(dy)\,\mu(dx) = \\ &= -n\int\int_{-\infty}^0\int_{y\leq t} y\,h'(x+t)\,\nu(dy)\,dt\,\mu(dx) + \\ &\quad + n\int\int_0^\infty\int_{y>t} y\,h'(x+t)\,\nu(dy)\,dt\,\mu(dx)\end{aligned}$$

Definirajmo funkcijo p: $\mathbb{R} \to \mathbb{R}$ po predpisu:

$$p(t) := \begin{cases} -n\int_{y\leq t} y\,\nu(dy) & t \leq 0 \\ n\int_{y>t} y\,\nu(dy) & t > 0 \end{cases}$$

Tedaj velja:

$$\mathbf{E}\big(Xh(X)\big) = \int\int_{-\infty}^{\infty} h'(x+t)\,p(t)\,dt\,\mu(dx) \tag{3.3.16}$$

Opomba. Odvod h' v z sicer ni definiran, vendar je vrednost izraza na desni neodvisna od tega, kako ga razširimo na celo realno os. Lahko se npr. dogovorimo, da postane h' z desne zvezna.

Funkcija p je očitno povsod nenegativna. Če v zgornjo enačbo vstavimo $f(x) = x$, dobimo $\int_{-\infty}^{\infty} p(t)\,dt = \mathbf{E}(X^2) = 1$. Funkcija p je torej gostota neke porazdelitve na $\mathbb{R}$. Razširimo zdaj verjetnostni prostor, na katerem so definirane slučajne spremenljivke $X_1, \ldots X_n$ tako, da bo na njem definirana še dodatna slučajna spremenljivka T, neodvisna od $X_1, \ldots X_n$ in s porazdelitvijo, ki ima za gostoto funkcijo p. Tedaj lahko enačbo (3.3.16) zapišemo tudi takole:

$$\mathbf{E}\big(Xh(X)\big) = \mathbf{E}\big(h'(X^* + T)\big)$$

Opomba. Porazdelitev slučajne spremenljivke $X^* + T$ je natančno porazdelitev slučajne spremenljivke X, transformirana na način, opisan v članku [11] kot 'zero-bias transformation' (glej formulo (3.1.7)). Tako obstaja možnost, da dobimo rezultat podobnega tipa, kot je Berry–Esseenov izrek, ne le pri neodvisnih slučajnih spremenljivkah, temveč tudi pri drugih problemih, pri katerih obstaja primerna slučajna spremenljivka s tako transformirano porazdelitvijo (glej članka [14] in [11]).

Vrnimo se zdaj k dokazu Berry–Esseenovega izreka. Po Chen–Steinovi metodi smo torej dobili:

$$\mathbf{E}\big(h'(X) - Xh(X)\big) = \mathbf{E}\big(h'(X^* + Y_n) - h'(X^* + T)\big)$$

Sledi:

$$d_{\mathrm{K}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) \le \sup_{z\in\mathbb{R}} \mathbf{E}\big(h_z'(X^* + Y_n) - h_z'(X^* + T)\big) \tag{3.3.17}$$

Funkcija h_z' je integral svojega odvoda na vsakem intervalu, ki ne vsebuje z. Na intervalih, ki vsebujejo z, pa je treba upoštevati še skok v z, ki smo ga izračunali v formuli (3.3.3). Velja torej:

$$h_z'(x + t) - h_z'(x) = \int_x^{x+t} h''(s)\,ds - I[x < z \le x + t] + I[x + t < z \le x] \tag{3.3.18}$$

Lema 3.3.3. *Za poljubne x, z in $t \in \mathbb{R}$ velja ocena:*

$$\left|\int_x^{x+t} h''(s)\,ds\right| \le |t| \left(\frac{1+\sqrt{3}}{2} + |x|\right)$$

Dokaz. Najprej pokažimo, da je:

$$\left|\int_x^{x+t} h_z''(s)\,ds\right| \le 1 \tag{3.3.19}$$

Pri dokazovanju te ocene lahko brez škode za splošnost privzamemo, da je $t > 0$. Če je $z \notin (x, x+t]$, velja kar $\int_x^{x+t} h_z''(s)\,ds = h_z'(x+t) - h_z'(x)$. Po točki (2) leme 3.3.2 števili $h_z'(x)$ in $h_z'(x+t)$ bodisi obe pripadata intervalu $[-1, 0]$ bodisi obe pripadata intervalu $[0, 1]$. Od tod pa takoj sledi (3.3.19). Če pa je $z \in (x, x+t]$, je $\int_x^{x+t} h_z''(s)\,ds = h_z'(x+t) - h_z'(x) + 1$. Toda v tem primeru število $h_z'(x + t)$ pripada intervalu $[-1, 0]$, število $h_z'(x)$ pa intervalu $[0, 1]$ in ocena (3.3.19) spet velja.

Integral funkcije h_z'' pa lahko ocenimo tudi po točki (3) leme 3.3.2. Velja:

$$\left|\int_x^{x+t} h_z''(s)\,ds\right| \leq \left|\int_x^{x+t} \big(1+|s|\big)\,ds\right| \leq \int_{|x|}^{|x|+|t|} \big(1+|s|\big)\,ds = |t|\big(1+|x|+\tfrac{1}{2}|t|\big)$$

Naj bo $|t| \leq \sqrt{3}-1$. Tedaj velja:

$$\left|\int_x^{x+t} h_z''(s)\,ds\right| \leq |t|\left(1+|x|+\frac{\sqrt{3}-1}{2}\right) = |t|\left(\frac{\sqrt{3}+1}{2}+|x|\right)$$

Če je $|t| \leq \sqrt{3}-1$, pa iz ocene (3.3.19) sledi:

$$\left|\int_x^{x+t} h_z''(s)\,ds\right| \leq 1 = (\sqrt{3}-1)\frac{\sqrt{3}+1}{2} \leq |t|\left(\frac{\sqrt{3}+1}{2}+|x|\right)$$

To pa je bilo treba dokazati. ■

Prepišimo zdaj formulo (3.3.18) malo drugače:

$$\begin{aligned} h_z'(x+y) - h_z'(x+t) &= \int_{x+t}^{x+y} h''(s)\,ds - I[x+y<z\leq x+t] + I[x+t<z\leq x+y] = \\ &= \int_x^{x+y} h''(s)\,ds - \int_x^{x+t} h''(s)\,ds - I[z-t\leq x<z-y]\,I[t>y] + \\ &\quad + I[z-y\leq x<z-t]\,I[t<y] \end{aligned}$$

Iz leme 3.3.3 zdaj sledi:

$$\begin{aligned} \big|h_z'(x+y) - h_z'(x+t)\big| \;\leq\; &\big(|y|+|t|\big)\big(\tfrac{1+\sqrt{3}}{2}+|x|\big) + I[z-t\leq x<z-y]\,I[t>y] + \\ &+ I[z-y\leq x<z-t]\,I[t<y] \end{aligned} \tag{3.3.20}$$

Opomba. V članku [14] je kot očitna navedena še nekoliko boljša ocena:

$$\begin{aligned} \big|h_z'(x+y) - h_z'(x+t)\big| \;\leq\; &\big(|y|+|t|\big)\big(1+|x|\big) + I[z-t\leq x<z-y]\,I[t>y] + \\ &+ I[z-y\leq x<z-t]\,I[t<y] \end{aligned} \tag{3.3.21}$$

Iz formule (3.3.17) in ocene (3.3.20) zdaj sledi:

$$\begin{aligned} d_{\mathrm{K}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) \leq \mathbf{E}\Big(&\big(|Y_n|+|T|\big)\big(\tfrac{1+\sqrt{3}}{2}+|X^*|\big) + \\ &+ I[z-T\leq X^*<z-Y_n]\,I[T>Y_n] + I[z-Y_n\leq X<z-T]\,I[T<Y_n]\Big) \end{aligned} \tag{3.3.22}$$

Najprej zaradi neodvisnosti velja:

$$\mathbf{E}\Big(\big(|Y_n|+|T|\big)\big(\tfrac{1+\sqrt{3}}{2}+|X^*|\big)\Big) = \big(\mathbf{E}(|Y_n|+\mathbf{E}(|T|)\big)\big(\tfrac{1+\sqrt{3}}{2}+\mathbf{E}(|X^*|)\big)$$

Označimo $\alpha := \mathbf{E}(|X_k|)$. Iz (3.1.1) sledi, da je $\alpha \le 1 \le \beta$. Velja $\mathbf{E}(|Y_n|) = n^{-1/2}\alpha$. Iz Jensenove neenakosti pa sledi:

$$\mathbf{E}(|X^*|)^2 \le \mathbf{E}\big((X^*)^2\big) = \mathrm{var}(X^*) = \frac{n-1}{n}$$

Torej je $\mathbf{E}(|X^*|) \le 1$. Nadalje je:

$$\begin{aligned}\mathbf{E}(|T|) = n\int_{-\infty}^{\infty}|t|\,p(t)\,dt &= -n\int_{-\infty}^{0} t\int_{y\le t} y\,\nu(dy)\,dt + n\int_0^{\infty} t\int_{y>t} y\,\nu(dy)\,dt = \\ &= -\,n\int_{y\le 0} y\int_y^0 t\,dt\,\nu(dy) + n\int_{y>0} y\int_0^y t\,dt\,\nu(dy) = \\ &= -\,\frac{n}{2}\int_{y\le 0} y^3\,\nu(dy) + \frac{n}{2}\int_{y>0} y^3\,\nu(dy) = \\ &= \frac{n}{2}\int |y|^3\,\nu(dy) = \frac{n}{2}\mathbf{E}\big(|Y_n|^3\big) = \frac{\beta}{2\sqrt{n}}\end{aligned} \tag{3.3.23}$$

Sledi:

$$\mathbf{E}\Big(\big(|Y_n| + |T|\big)\big(\tfrac{1+\sqrt{3}}{2} + |X^*|\big)\Big) \le \frac{3+\sqrt{3}}{4\sqrt{n}}(2\alpha+\beta) \tag{3.3.24}$$

Ocenili smo torej prvi člen na desni strani ocene (3.3.22). Druga dva člena pa najprej zapišemo kot dvakratni integral. Dobimo:

$$\begin{aligned}&\mathbf{E}\Big(I[z-T \le X^* < z-Y_n]\,I[T>Y_n] + I[z-Y_n \le X < z-T]\,I[T<Y_n]\Big) = \\ &= \int_{t>y} \mathbf{P}[z-t \le X^* < z-y]\mathcal{L}(Y_n,T)(dy\times dt) + \\ &\quad + \int_{t<y} \mathbf{P}[z-y \le X^* < z-t]\mathcal{L}(Y_n,T)(dy\times dt)\end{aligned} \tag{3.3.25}$$

Oceniti je torej treba verjetnost, da X^* pripada danemu intervalu $(a,b]$. To je ključna ocena pri dokazu Berry–Esseenovega izreka, saj nam prav ta premosti skok odvoda h_z' v točki z.

Lema 3.3.4. *Za poljubni realni števili a in b, za kateri je $a<b$, velja ocena:*

$$\mathbf{P}[a \le X^* < b] \le b - a + \frac{2\beta}{\sqrt{n}}$$

Dokaz. Najprej po neenačbi Markova ocenimo:

$$\mathbf{P}\left[|T| \le \tfrac{\beta}{\sqrt{n}}\right] = 1 - \mathbf{P}\left[|T| \le \tfrac{\beta}{\sqrt{n}}\right] \ge 1 - \tfrac{\sqrt{n}}{\beta}\mathbf{E}(|T|) = \tfrac{1}{2}$$

Upoštevali smo formulo (3.3.23). Sledi:

$$\begin{aligned}\mathbf{P}[a \le X^* < b] &\le 2\mathbf{P}[a \le X^* < b]\,\mathbf{P}\left[|T| \le \tfrac{\beta}{\sqrt{n}}\right] = 2\mathbf{P}\left[a \le X^* < b, |T| \le \tfrac{\beta}{\sqrt{n}}\right] \le \\ &\le 2\mathbf{P}\left[a - \tfrac{\beta}{\sqrt{n}} \le X^* + T < a + \tfrac{\beta}{\sqrt{n}}\right] = 2\mathbf{E}\big(g(X^*+T)\big)\end{aligned}$$

kjer je $g(x) = I\left[a - \frac{\beta}{\sqrt{n}} \le x < b + \frac{\beta}{\sqrt{n}}\right]$. Naj bo:

$$h(x) := \int_{\frac{a+b}{2}}^{x} g(t)\,dt = \begin{cases} -\frac{1}{2}(b-a) - \frac{\beta}{\sqrt{n}} & x \le a - \frac{\beta}{\sqrt{n}} \\ x - \frac{1}{2}(a+b) & a - \frac{\beta}{\sqrt{n}} \le x \le b + \frac{\beta}{\sqrt{n}} \\ \frac{1}{2}(b-a) + \frac{\beta}{\sqrt{n}} & x \ge b + \frac{\beta}{\sqrt{n}} \end{cases}$$

Tedaj po formuli (3.3.16) velja:

$$\begin{aligned} \mathbf{P}[a \le X^* \le b] &\le 2\mathbf{E}\big(h'(X^* + T)\big) = 2\mathbf{E}\big(Xh(X)\big) \le 2\mathbf{E}\big(|Xh(X)|\big) \le \\ &\le \big(b - a + \tfrac{2\beta}{\sqrt{n}}\big)\mathbf{E}(|X|) \le b - a + \tfrac{2\beta}{\sqrt{n}} \end{aligned}$$

Upoštevali smo $\mathbf{E}(|X|)^2 \le \mathbf{E}(X^2) = \mathrm{var}(X) = 1$. Lema je tako dokazana. ■

Iz zgornje leme in ocene (3.3.25) zdaj dobimo:

$$\begin{aligned} &\mathbf{E}\Big(I[z - T \le X^* < z - Y_n]\, I[T > Y_n] + I[z - Y_n \le X < z - T]\, I[T < Y_n]\Big) \le \\ &\le \mathbf{E}\Big(I[T > Y_n]\big(T - Y_n + \tfrac{2\beta}{\sqrt{n}}\big) + I[T < Y_n]\big(Y_n - T + \tfrac{2\beta}{\sqrt{n}}\big)\Big) = \\ &= \frac{2\beta}{\sqrt{n}} + \mathbf{E}\big(|Y_n - T|\big) \le \frac{2\beta}{\sqrt{n}} + \mathbf{E}(|Y_n|) + \mathbf{E}(|T|) \le \\ &\le \frac{2\beta}{\sqrt{n}} + \frac{\alpha}{\sqrt{n}} + \frac{\beta}{2\sqrt{n}} = \frac{2\alpha + 5\beta}{2\sqrt{n}} \end{aligned}$$

Iz te ocene ter ocen (3.3.22) in (3.3.24) zdaj končno sledi:

$$\begin{aligned} d_{\mathrm{K}}\big(\mathcal{L}(X), \mathrm{N}(0,1)\big) &\le \frac{3 + \sqrt{3}}{4\sqrt{n}}(2\alpha + \beta) + \frac{2\alpha + 5\beta}{2\sqrt{n}} = \frac{(10 + 2\sqrt{3})\alpha + (13 + \sqrt{3})\beta}{4\sqrt{n}} \le \\ &\le \frac{23 + 3\sqrt{3}}{4} \frac{\beta}{\sqrt{n}} < 7.05 \frac{\beta}{\sqrt{n}} \end{aligned}$$

Tako je Berry–Esseenov izrek (do konstante natančno) dokazan.

Opomba. V članku [14], kjer je namesto ocene (3.3.20) uporabljena ocena (3.3.21), v zgornji oceni namesto konstante 7.05 stoji konstanta 6.5. Stein v svoji knjigi [24] navede še en dokaz Berry–Esseenovega izreka in dobi konstanto 6 (pri čemer pripomni, da bi se dala s pazljivejšim ocenjevanjem še izboljšati). Vse te konstante pa so daleč od že znane ocene s konstanto 0.8. Dobili smo sicer pravo hitrost konvergence, vendar pa je tu bolj kot rezultat pomembna ideja, saj se jo da uporabiti tudi pri neenako porazdeljenih in celo odvisnih slučajnih spremenljivkah.

4.

Večrazsežna normalna aproksimacija

V tem poglavju bomo rezultate prejšnjega poglavja posplošili na več dimenzij. Aproksimirali bomo torej porazdelitev slučajnega vektorja z vrednostmi v $\mathbb{R}^p$, in sicer s standardno večrazsežno normalno porazdelitvijo $\gamma_p := \mathrm{N}(0, I)$, kjer je I identiteta na $\mathbb{R}^p$.

Novost pri večrazsežni normalni aproksimaciji je rešitev Steinove enačbe. Tako pri Poissonovi kot tudi pri enorazsežni normalni aproksimaciji smo se lahko izmazali brez izreka 1.0.1, saj smo npr. pri enorazsežni normalni aproksimaciji v Steinovo enačbo:

$$g''(x) - x\,g'(x) = f(x) - \int f\,d\gamma_1 \tag{4.0.1}$$

lahko vpeljali substitucijo $h = g'$. Tudi pri večrazsežni normalni aproksimaciji bomo za osnovo vzeli Ornstein–Uhlenbeckov proces (glej dodatek B, strani 126–135) in dejstvo, da je njegova invariantna porazdelitev standardna normalna (glej stran 138). Steinova enačba se bo tako glasila:

$$(\Delta g)(x) - x \cdot (\nabla g)(x) = f(x) - \int f\,d\gamma_p$$

Te enačbe pa se ne da rešiti tako lepo kot pri enorazsežni normalni aproksimaciji. V naslednjem razdelku se bomo oprli na izrek 1.0.1 in pokazali, da ima enačba rešitev oblike:

$$g = -\int_0^\infty \left(P_t f - \int f\,d\gamma_p\right) dt$$

kjer so P_t elementi polgrupe, ki pripada Ornstein–Uhlenbeckovemu procesu, t. j.:

$$(P_t f)(x) = \int f\Big(e^{-t}x + \sqrt{1 - e^{-2t}}\,y\Big)\gamma_p(dy)$$

Nato bomo s parcialnimi odvodi funkcije f ocenili parcialne odvode funkcije g. Brž ko bomo torej za dani slučajni vektor X znali izraz:

$$\mathbf{E}\big((Gg)(X)\big) = \mathbf{E}\big((\Delta g)(X) - X \cdot (\nabla g)(X)\big)$$

oceniti s parcialnimi odvodi funkcije g, bomo znali tudi izraz $\int f\, d\mathcal{L}(X) - \int f\, d\gamma_p$ oceniti s parcialnimi odvodi funkcije f istih redov.

Pri enorazsežni normalni aproksimaciji smo videli, da se izraz $\mathbf{E}\big((Gg)(X)\big)$ za neodvisne in enako porazdeljene slučajne spremenljivke oceni z $M_3(g)$ oziroma s tretjimi parcialnimi odvodi. Podobne ocene nastopajo tudi pri večrazsežni normalni aproksimaciji. Pri enorazsežni normalni aproksimaciji smo videli, da ima rešitev g Steinove enačbe (4.0.1) omejen tretji odvod. Žal dokaz tega dejstva v večrazsežnem primeru povsem odpove, saj temelji na rešitvi, dobljeni s substitucijo $h = g'$, ki v večrazsežnem primeru ni izvedljiva. Vprašanje, kako je s tretjimi parcialnimi odvodi funkcije g pri večrazsežni normalni aproksimaciji, bomo pustili odprto. S tem pa seveda odpove tudi pot, ki smo jo ubrali pri ocenjevanju v Wassersteinovi metriki. Le-to pa se da vseeno dobiti, in sicer tako, da funkcijo f najprej malo zgladimo in šele nato ocenili po Chen–Steinovi metodi. Idejo povzemamo iz članka [13], v katerem je naveden celo rezultat Berry–Esseenovega tipa (testne funkcije so indikatorji merljivih konveksnih množic).

Seveda tudi večrazsežna normalna aproksimacija ne deluje le za neodvisne in enako porazdeljene slučajne spremenljivke, temveč celo za odvisne. Več o tem si lahko bralec pogleda v člankih [11] in [12]. Sami postopki ocenjevanja so podobni kot v enodimenzionalnem primeru, bistveno različno je le ocenjevanje rešitve Steinove enačbe, ki pa se tako ali tako naredi enkrat za vselej.

4.1 Rešitev Steinove enačbe

Naj bo $(P_t)_{t\in[0,\infty)}$ krepko zvezna operatorska polgrupa z generatorjem G. Izrek 1.0.1 nam reši Steinovo enačbo:

$$Gg = f - \int f\, d\mu$$

za vsako funkcijo f, za katero funkcije $P_t f$ dovolj hitro konvergirajo proti $\int f\, d\mu$. V našem primeru moramo to narediti za Ornstein–Uhlenbeckov proces na $\mathbb{R}^p$ in standardno normalno porazdelitev $\mu = \gamma_p$. Če hočemo zadostiti pogojem izreka 1.0.1, morajo funkcije $P_{t_n} f$ konvergirati proti $\int f\, d\gamma_p$ za vsaj eno zaporedje t_n, ki gre proti neskončno. Konvergenca je seveda mišljena v normi prostora funkcij. Naravni prostor funkcij, na katerem gledamo Feller–Dynkinove polgrupe, je seveda $\mathcal{C}_0(\mathbb{R}^p)$. Toda ta prostor za našo nalogo ni primeren, saj konstanta $\int f\, d\gamma_p$ navadno ne pripada temu prostoru. Tudi če prostor $\mathcal{C}_0(\mathbb{R}^p)$ dopolnimo s konstantami in dobimo prostor vseh zveznih funkcij na kompaktifikaciji prostora $\mathbb{R}^p$ z eno točko, t. j. na sferi S^p, nam vse to prav nič ne pomaga, saj je prostor $\mathcal{C}_0(\mathbb{R}^p)$ zaprt v tem večjem prostoru. Sama razširitev prostora $\mathcal{C}_0(\mathbb{R}^p)$ torej ne bo dovolj, popraviti bo treba tudi normo.

Kar bomo naredili, je tole: prostor $\mathcal{C}_0(\mathbb{R}^p)$ bomo preslikali s primerno zvezno funkcijo $\psi\colon \mathbb{R}^p \to (0,\infty)$. Definirali bomo torej:

$$L := \big\{\psi f \;\big|\; f \in \mathcal{C}_0(\mathbb{R}^p)\big\} = \Big\{g \in \mathcal{C}(\mathbb{R}^p) \;\Big|\; \tfrac{1}{\psi}\, g \in \mathcal{C}_0(\mathbb{R}^p)\Big\}$$

Očitno je L vektorski prostor. Na njem definirajmo normo:

$$\|g\|_L := \big\|\tfrac{1}{\psi}\, g\big\|_\infty$$

kjer je $\|\cdot\|_\infty$ običajna supremum norma. Tako opremljen prostor L je izometrično izomorfen prostoru $\mathcal{C}_0(\mathbb{R}^p)$ in je zato Banachov prostor.

Če želimo, da naš prostor L vsebuje konstante, bo moralo veljati $\lim_{\|x\|\to\infty} \psi(x) = \infty$. Izbrali bomo:

$$\psi(x) = e^{c\|x\|}, \quad c > 0$$

Površno povedano, Steinovo enačbo bomo rešili za funkcije, ki imajo največ eksponentno rast. No, v resnici bodo morale biti funkcije še dvakrat zvezno odvedljive in tudi njihovi prvi in drugi parcialni odvodi bodo smeli imeti največ eksponentno rast.

Opomba. Ideja konstrukcije prostora L je povzeta iz članka [4], kjer avtor vzame funkcijo $\psi(x) := 1 + \|x\|^3$. Članek sam se ukvarja z neskončnorazsežno aproksimacijo: namesto z normalno porazdelitvijo aproksimiramo z Brownovim gibanjem. Pri konstrukciji prostora L pa avtor naredi majhen spodrsljaj: namesto prostora $\mathcal{C}_0(\mathbb{R}^p)$ pomnoži s funkcijo ψ kar cel prostor $\mathcal{C}(\mathbb{R}^p)$. Toda v tem primeru operatorska polgrupa Ornstein–Uhlenbeckovega procesa (glej spodaj) ni več krepko zvezna (avtor privzeme krepko zveznost kot očitno). Protiprimer je funkcija $f(x) := \psi(|x_1|)\sin(x_1^2)$. No, v resnici avtor namesto prostora $\mathbb{R}^p$ gleda prostor določenih funkcij na $[0, 1]$ s supremum normo, toda v tem primeru namesto projekcije $x \mapsto x_1$ vzamemo projekcijo $u \mapsto u(0)$ in prav tako dobimo protiprimer.

Oglejmo si, kako se na prostoru L obnaša operatorska polgrupa Ornstein–Uhlenbeckovega procesa:

$$(P_t f)(x) = \int f\Big(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\Big)\gamma_p(dy)$$

Najprej moramo preveriti, da operatorji P_t sploh slikajo L v L. Naj bo $f \in L$. Dokazati moramo:

$$\lim_{\|x\|\to\infty} e^{-c\|x\|} \int f\Big(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\Big)\gamma_p(dy) = 0$$

Ker velja:

$$\begin{aligned} e^{-c\|x\|} \left| f\left(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\right)\right| &\le e^{-c\|x\|}\|f\|_L e^{c\|e^{-t}x + \sqrt{1-e^{-2t}}y\|} \\ &\le \|f\|_L e^{-c\|x\|} e^{c\|x\|} e^{c\|y\|} = \|f\|_L e^{c\|y\|} \end{aligned}$$

in je izraz na desni v $L^1(\gamma_p(dy))$, lahko uporabimo izrek o dominirani konvergenci. In res za vsak y velja:

$$\begin{aligned} &\lim_{\|x\|\to\infty} e^{-c\|x\|} \left| f\left(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\right)\right| \le \\ &\quad \le \lim_{\|x\|\to\infty} e^{-c\|e^{-t}x + \sqrt{1-e^{-2t}}y\|} \left| f\left(e^{-t}x + \sqrt{1 - e^{-2t}} y\right)\right| e^{c(e^{-t}-1)\|x\|} e^{c\|y\|} = 0 \end{aligned}$$

torej operatorji P_t res slikajo L v L. Preverimo, da so operatorji P_t omejeni. Velja:

$$\begin{aligned} e^{-c\|x\|}|(P_t f)(x)| &\le e^{-c\|x\|} \int \left| f\left(e^{-tx} + \sqrt{1 - e^{-2t}}\, y\right)\right| \gamma_p(dy) \le \\ &\le e^{-c\|x\|}\|f\|_L \int e^{c\|x\|} e^{c\|y\|}\gamma_p(dy) = \|f\|_L \int e^{c\|y\|}\gamma_p(dy) \end{aligned}$$

Sledi $\|P_t\|_L \le \int e^{\|y\|}\,\gamma_p(dy)$, torej so operatorji P_t enakomerno omejeni.

Operatorji P_t tudi na novem prostoru tvorijo polgrupo. Vemo namreč, da enakost $P_{t+s}f = P_tP_sf$ velja na prostoru $\mathcal{C}_0(\mathbb{R}^p)$, toda ta prostor je gost v L. To pa pomeni, da enakost velja za vsak $f \in L$.

Na 'naravnem' prostoru $\mathcal{C}_0(\mathbb{R}^p)$ je operatorska polgrupa $(P_t)_{t\in[0,\infty)}$ krepko zvezna, saj je dobljena iz Feller–Dynkinovega procesa. Pri tem je ključen izrek B.8.4. Nastane vprašanje, kako je s krepko zveznostjo na L. Tudi tu nam pomega izrek B.8.4.

Trditev 4.1.1. *Operatorska polgrupa $(P_t)_{t\in[0,\infty)}$ je tudi na prostoru L zvezna (za vsak $c > 0$).*

Dokaz. Iz izreka o dominirani konvergenci takoj sledi, da za vsak $f \in L$ in vsak $x \in \mathbb{R}^p$ velja $\lim_{t\downarrow 0}(P_tf)(x) = f(x)$. Definirajmo zdaj operatorje P'_t na prostoru $\mathcal{C}_0(\mathbb{R}^p)$ po predpisu $P'_tf := \frac{1}{\psi}P_t(\psi f)$. Tudi ti operatorji so enakomerno omejeni in za vsak $f \in \mathcal{C}_0(\mathbb{R}^p)$ in vsak $x \in \mathbb{R}^p$ velja $\lim_{t\downarrow 0}(P'_tf)(x) = f(x)$. Po izreku B.8.4 je potem operatorska polgrupa $(P'_t)_{t\in[0,\infty)}$ krepko zvezna, potem pa to velja tudi za operatorsko polgrupo $(P_t)_{t\in[0,\infty)}$. ■

Opomba. Operatorji naše polgrupe so sicer enakomerno omejeni, vendar pa niso kontrakcije. Vzemimo namreč $f(x) := 1 + c\|x\|$. Ta funkcija je v L in velja $\|f\|_L = 1$, vendar pa je $(P_tf)(0) > 1$ za vsak $t > 0$ in zato tudi $\|P_tf\|_L > 1$.

Izračunajmo generator naše operatorske polgrupe! To bomo storili podobno kot pri izpeljavi generatorja na $\mathcal{C}_0(\mathbb{R}^p)$ (glej strani 131–135). Naj bo torej $f \in L \cap \mathcal{C}^2(\mathbb{R}^p)$ in naj bodo vsi drugi parcialni odvodi funkcije f v L. Razvoj v Taylorjevo vrsto nam podobno kot pri izpeljavi na $\mathcal{C}_0(\mathbb{R}^p)$ da:

$$
\begin{aligned}
f\left(e^{-t}x + \sqrt{1-e^{-2t}}\,y\right) &= f(e^{-t}x) + \sqrt{1-e^{-2t}}\sum_{i=1}^p \frac{\partial f}{\partial x_i}(e^{-t}x)\,y_i + \\
&\quad + \frac{1-e^{-2t}}{2}\sum_{i=1}^p\sum_{j=1}^p \frac{\partial^2 f}{\partial x_i\,\partial x_j}(e^{-t}x)\,y_i\,y_j + R(x,y,t)
\end{aligned}
$$

kjer je:

$$
\begin{aligned}
|R(x,y,t)| &\le \frac{1-e^{-2t}}{2}\rho\big(x, \sqrt{1-e^{-2t}}\|y\|\big)\sum_{i=1}^p\sum_{j=1}^p |y_iy_j| \\
\rho(x,r) &:= \sup_{\substack{\|y-x\|\le r\\ 1\le i,j\le p}} \left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x) - \frac{\partial^2 f}{\partial x_i\,\partial x_j}(y)\right|
\end{aligned}
$$

Ko integriramo po $\gamma_p(dy)$, dobimo:

$$
(P_tf)(x) = f(e^{-t}x) + \frac{1-e^{-2t}}{2}(\Delta f)(e^{-t}x) + \int R(x,y,t)\,\gamma_p(dy)
$$

Zgornji izraz lahko gledamo tudi kot preslikavo iz $[0,\infty)$ v $\mathcal{C}_0(\mathbb{R}^p)$ in izračunati moramo desni odvod te preslikave pri $t = 0$. Pišimo:

$$
P_tf = u_1(t) + \frac{1-e^{-2t}}{2t}\big(u_2(t) + u_3(t) + u_4(t)\big)
$$

kjer je:

$$u_1(t)(x) := f(e^{-t}x), \quad u_2(t)(x) := t(\Delta f)(x), \quad u_3(t)(x) := t\big((\Delta f)(e^{-t}x) - (\Delta f)(x)\big),$$
$$u_4(t)(x) := \frac{2t}{1-e^{-2t}} \int R(x,y,t)\,\gamma_p(dy), \quad u_4(0) := 0$$

Dovolj je preveriti, da so vse štiri funkcije z desne odvedljive v 0. Tudi pri odvajanju preslikav z vrednostmi v Banachovem prostoru namreč velja pravilo za odvajanje produkta: če je α odvedljiva realna funkcija, u pa odvedljiva funkcija z vrednostmi v Banachovem prostoru, je odvedljiva tudi funkcija αu in velja $(\alpha u)' = \alpha' u + \alpha u'$. V našem primeru je:

$$\alpha(t) = \begin{cases} \dfrac{1-e^{-2t}}{2t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

Funkcija α je odvedljiva (saj je celo analitična). Ker je $u_2(0) = u_3(0) = u_4(0) = 0$, velja:

$$Gf = \left(\frac{d}{dt}\right)^+ \Bigg|_{t=0} P_t f = u_1'(0) + u_2'(0) + u_3'(0) + u_4'(0)$$

brž ko imajo vse preslikave u_1, u_2, u_3 in u_4 desni odvod v 0 (seveda v prostoru L). Za u_1, u_2 in u_3 bi se dalo pokazati, da imajo tudi obojestranske odvode, vendar pa tega ne bomo potrebovali. Če zapišemo:

$$u_1(t) = v(e^{-t}), \quad v(s)(x) := f(sx)$$

po verižnem pravilu velja $u_1'(0) = -v'(1)$, kjer je pri u mišljen desni, pri v pa levi odvod. Velja:

$$v(1-h)(x) = f((1-h)x) = f(x) + h \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}((1-\vartheta h)x)\, x_i =$$
$$= v(1) - hx \cdot (\nabla f)((1-\vartheta h)x)$$

kjer je $0 \leq \vartheta \leq 1$. Domnevamo torej, da je:

$$v'(1)(x) = x \cdot (\nabla f)(x) \tag{4.1.1}$$

Ta funkcija pa bo morala biti v L. Za funkcijo f bomo torej privzeli še, da je:

$$\lim_{\|x\|\to\infty} e^{-c\|x\|} x \cdot (\nabla f)(x) = 0$$

Dokažimo zdaj, da je v res z leve odvedljiva v 1. Naj bo $\varepsilon > 0$. Obstaja tak $r > 0$, da za vsak $y \in \mathbb{R}^p$, za katerega je $\|y\| > r$, velja $|y \cdot (\nabla f)(y)| < \varepsilon/3$. Naj bo zdaj $\|x\| > 2r$ in $0 \leq h < \frac{1}{2}$. Tedaj je gotovo $\|(1+\vartheta h)x\| > r$. Sledi:

$$e^{-c\|x\|} \left| \frac{v(1-h)(x) - v(1)(x)}{-h} - x \cdot (\nabla f)(x) \right| =$$
$$= e^{-c\|x\|} \left| x \cdot (\nabla f)((1-\vartheta h)x) - x \cdot (\nabla f)(x) \right| \leq$$
$$\leq \frac{e^{-c(1-\vartheta h)\|x\|}}{1-\vartheta h} |(1-\vartheta h)x \cdot (\nabla f)((1-\vartheta h)x)| + e^{-c\|x\|} |x \cdot (\nabla f)(x)| < \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

Na množici $\{y \mid \|y\| \leq 2r\}$ pa je preslikava ∇f enakomerno zvezna. Obstaja torej tak $\delta > 0$, da za poljubna y in z z normo največ $2r$ in oddaljena za manj kot δ velja $\|(\nabla f)(y) - (\nabla f)(z)\| < \frac{\varepsilon}{2r}$. Naj bo zdaj $\|x\| \leq 2r$ in $0 \leq h < \min\left\{\frac{1}{2}, \frac{\delta}{2r}\right\}$. Tedaj velja:

$$e^{-c\|x\|}\left|\frac{v(1+h)(x) - v(1)(x)}{-h} - x \cdot (\nabla f)(x)\right| \leq \left|x \cdot (\nabla f)((1+\vartheta h)x) - x \cdot (\nabla f)(x)\right| \leq$$
$$\leq \|x\| \, \|(\nabla f)((1+\vartheta h)x) - (\nabla f)(x)\| < 2r \, \frac{\varepsilon}{2r} = \varepsilon$$

Dokazali smo, da, brž ko je $0 \leq h < \min\left\{\frac{1}{2}, \frac{\delta}{2r}\right\}$, velja:

$$\left\|\frac{v(1-h) - v(1)}{-h} - \big(x \mapsto x \cdot (\nabla f)(x)\big)\right\|_L < \varepsilon$$

To pa pomeni, da je preslikava v z leve odvedljiva v 1 in da velja (4.1.1), torej:

$$u_1'(0)(x) = -x \cdot (\nabla f)(x)$$

Očitno je $u_2'(0) = \Delta f$. Pokažimo še, da je $u_3'(0) = u_4'(0) = 0$, kjer seveda mislimo desna odvoda. Ker je $u_3(0) = u_4(0) = 0$, bo dovolj pokazati, da je $\lim_{t \downarrow 0} \frac{\|u_3(t)\|_L}{t} = \lim_{t \downarrow 0} \frac{\|u_4(t)\|_L}{t} = 0$. Najprej je:

$$\frac{|u_3(t)(x)|}{t} = |(\Delta f)(e^{-t}x) - (\Delta f)(x)|$$

Naj bo $\varepsilon > 0$. Ker vsi drugi parcialni odvodi funkcije f pripadajo L, obstaja tak $r > 0$, da za vsak $y \in \mathbb{R}^p$, za katerega je $\|y\| > r$, velja $e^{-c\|y\|}|(\Delta f)(y)| < \varepsilon/2$. Naj bo zdaj $\|x\| > er$ in $0 \leq t < 1$. Tedaj je $\|e^{-t}x\| > r$, zato velja:

$$e^{-c\|x\|}|(\Delta f)(e^{-t}x) - (\Delta f)(x)| \leq e^{-c\|x\|}|(\Delta f)(e^{-t}x)| + e^{-c\|x\|}|(\Delta f)(x)| <$$
$$< e^{-c\|x\|}e^{ce^{-t}\|x\|}\frac{\varepsilon}{2} + \frac{\varepsilon}{2} \leq \varepsilon$$

Na množici $\{y \mid \|y\| \leq er\}$ pa je funkcija Δf enakomerno zvezna. Obstaja torej tak $\delta > 0$, da za poljubna y in z z normo največ er in oddaljena za manj kot δ velja $|(\Delta f)(y) - (\Delta f)(z)| < \varepsilon$. Brez škode za splošnost lahko privzamemo, da je $\delta < er$. Naj bo zdaj $\|x\| \leq er$ in $0 \leq t < -\log\left(1 - \frac{\delta}{er}\right)$. Tedaj je gotovo $\|e^{-t}x\| \leq er$ in $\|(1 - e^{-t})x\| < \delta$. Sledi:

$$e^{-c\|x\|}|(\Delta f)(e^{-t}x) - (\Delta f)(x)| \leq |(\Delta f)(e^{-t}x) - (\Delta f)(x)| < \varepsilon$$

Dokazali smo, da, brž ko je $0 \leq t < \min\left\{1, -\log\left(1 - \frac{\delta}{er}\right)\right\}$, velja $\frac{\|u_3(t)\|_L}{t} < \varepsilon$. To pa pomeni, da je res $u_3'(0) = 0$.

Dokažimo še, da je tudi $u_4'(0) = 0$. Definirajmo:

$$\varphi(r) := \sup_{x \in \mathbb{R}}\big(e^{-c\|x\|}\rho(x, r)\big)$$

Tedaj velja:

$$\frac{\|u_4(t)\|_L}{t} = \frac{2}{1 - e^{-2t}} \sup_x \left(e^{-c\|x\|} \int R(x, y, t)\, \gamma_p(dy)\right) \leq$$
$$\leq \int \varphi\big(\sqrt{1 - e^{-2t}}\|y\|\big) \sum_{i=1}^p \sum_{j=1}^p |y_i y_j| \, \gamma_p(dy)$$

Dokazati moramo, da gre izraz na desni proti 0, ko gre t proti 0. Seveda bomo uporabili izrek o dominirani konvergenci. Integrand je omejen s $\varphi(\|y\|)\sum_{i,j}|y_i y_j|$. Pokažimo, da je ta funkcija v $L^1(\gamma_p(dy))$. Ker so drugi parcialni odvodi v L, obstaja tak $M > 0$, da za vsak x ter poljubna i in j velja:

$$e^{-c\|x\|}\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right| \leq M$$

Naj bo zdaj $\|z - x\| \leq r$. Tedaj velja:

$$\begin{aligned} e^{-c\|x\|}\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(z) - \frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right| &\leq e^{-c\|x\|}\left(\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right| + \left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(z)\right|\right) \leq \\ &\leq M\big(1 + e^{c(\|z\|-\|x\|)}\big) \leq M\big(1 + e^{cr}\big) \end{aligned}$$

Od tod pa sledi, da je $\varphi(r) \leq M(1 + e^{cr})$, od tod pa sledi, da je res $\int \varphi(\|y\|)\sum_{i,j}|y_i y_j| \times \times \gamma_p(dy) < \infty$. To pa pomeni, da lahko uporabimo izrek o dominirani konvergenci. Dovolj je torej dokazati:

$$\lim_{t\downarrow 0}\varphi(t) = 0$$

Naj bo $\varepsilon > 0$. Obstaja tak $r > 0$, da za vsak x, za katerega je $\|x\| > r$, velja:

$$e^{-c\|x\|}\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right| < \frac{\varepsilon}{1 + e^c}$$

Naj bo zdaj $\|x\| > r + 1$ in $\|y - x\| < 1$. Tedaj je gotovo $\|y\| < r + 2$. Sledi:

$$\begin{aligned} e^{-c\|x\|}\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(y) - \frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right| &\leq \left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right| + e^{c(\|y\|-\|x\|)}\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(y)\right| < \\ &< \frac{\varepsilon}{1 + e^c} + e^c\,\frac{e^c}{1 + e^c} = \varepsilon \end{aligned}$$

Brž ko je torej $\|x\| \geq r + 1$ in $0 \leq t < 1$, je $e^{-c\|x\|}\rho(x,t) < \varepsilon$.

Na množici $\{x \mid \|x\| \leq r+2\}$ pa so drugi parcialni odvodi enakomerno zvezni. Obstaja torej tak $\delta > 0$, da za poljubna i in j ter poljubna x in y, oddaljena za manj kot δ in z normo največ $r + 2$, velja:

$$\left|\frac{\partial^2 f}{\partial x_i\,\partial x_j}(y) - \frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right| < \varepsilon$$

Brž ko je torej $\|x\| \leq r + 1$ in $0 \leq t < \delta$, velja:

$$e^{-c\|x\|}\rho(x,t) \leq \rho(x,t) \leq \varepsilon$$

Brž ko je torej $0 \leq t < \min\{1, \delta\}$, za vsak $x \in \mathbb{R}^p$ velja $e^{-c\|x\|}\rho(x,t) \leq \varepsilon$. Od tod pa sledi tudi, da je $\varphi(t) \leq \varepsilon$, torej je res $\lim_{t\downarrow 0}\varphi(t) = 0$, od tod pa sledi, da je tudi $u_4'(0) = 0$. Dokazali smo naslednji izrek.

Izrek 4.1.2. *Naj bo $(P_t)_{t\in[0,\infty)}$ operatorska polgrupa, ki pripada Ornstein–Uhlenbeckovemu procesu, in naj bo:*

$$D := \left\{ f \in L \cap \mathcal{C}^2(\mathbb{R}^p) \;\middle|\; (\forall i, j) \tfrac{\partial^2 f}{\partial x_i \partial x_j} \in L, \big(x \mapsto x \cdot (\nabla f)(x)\big) \in L \right\} \tag{4.1.2}$$

Naj bo G generator operatorske polgrupe $(P_t)_{t\in[0,\infty)}$. Tedaj je $D \subset \mathcal{D}(G)$ in za vsak $f \in D$ velja:

$$(Gf)(x) = (\Delta f)(x) - x \cdot (\nabla f)(x) \tag{4.1.3}$$

Tako kot na 'naravnem' prostoru $\mathcal{C}_0(\mathbb{R}^p)$ se je tudi tu smiselno vprašati, ali je G zaprtje svoje zožitve na D. Odgovor je seveda pritrdilen, dokaz pa je čisto enak kot na prostoru $\mathcal{C}_0(\mathbb{R}^p)$ in ga bomo zato opustili (glej dodatek B). Tega dejstva niti ne bomo potrebovali.

Pokažimo zdaj še konvergenco proti ravnovesni porazdelitvi, ki jo zahteva izrek 1.0.1 in zaradi katere sploh obravnavamo operatorje P_t na prostoru L. Za dano funkcijo f moramo torej preveriti, kako hitro elementi $P_t f$ konvergirajo proti $\int f\, d\gamma_p$, ko gre t proti neskončno. Velja:

$$\begin{aligned}
\int f\, d\gamma_p &= \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} f(y)\, e^{-\frac{1}{2}\|y\|^2} dy \\
(P_t f)(x) &= \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} f\left(e^{-t}x + \sqrt{1-e^{-2t}}\, y\right) e^{-\frac{1}{2}\|z\|^2} dz = \\
&= \frac{1}{(2\pi)^{p/2}(1-e^{-2t})^{p/2}} \int f(y)\, e^{-\frac{1}{2}\frac{\|y-e^{-t}x\|^2}{1-e^{-2t}}} dy
\end{aligned}$$

Če torej definiramo:

$$g(x, y, s) := \frac{1}{(2\pi)^{p/2}(1-s^2)^{p/2}} e^{-\frac{1}{2}\frac{\|y-sx\|^2}{1-s^2}}$$

velja:

$$(P_t f)(x) - \int f\, d\gamma_p = \int_{\mathbb{R}^p} f(y)\big(g(x, y, e^{-t}) - g(x, y, 0)\big) dy$$

Po Lagrangeovem izreku velja:

$$g(x, y, e^{-t}) - g(x, y, 0) = \frac{\partial g}{\partial s}(x, y, s), \quad 0 \le s \le e^{-t}$$

Po krajšem računu dobimo:

$$\frac{\partial g}{\partial s}(x, y, s) = \frac{1}{(2\pi)^{p/2}(1-s^2)^{p/2+2}} e^{-\frac{1}{2}\frac{\|y-sx\|^2}{1-s^2}} \left(ps(1-s^2) + (1-s^2)(y-sx)\cdot x - s\|y-sx\|^2\right)$$

Naj bo $t \ge 1$. Tedaj je tudi $0 \le s \le 1/e$. Sledi:

$$\left|\frac{\partial g}{\partial s}(x, y, s)\right| \le \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}\|y-sx\|^2} \left(\frac{p}{e(1-e^{-2})^{p/2+1}} + \frac{\|y-sx\|\,\|x\|}{(1-e^{-2})^{p/2+1}} + \frac{\|y-sx\|^2}{e(1-e^{-2})^{p/2+2}}\right)$$

Če definiramo:

$$M_0 := \frac{p}{(2\pi)^{p/2} e(1-e^{-2})^{p/2+1}}, \quad M_1 := \frac{1}{(2\pi)^{p/2}(1-e^{-2})^{p/2+1}} \max_{u \geq 0}\left(ue^{-\frac{1}{4}u^2}\right),$$
$$M_2 := \frac{1}{(2\pi)^{p/2} e(1-e^{-2})^{p/2+2}} \max_{u \geq 0}\left(u^2 e^{-\frac{1}{4}u^2}\right)$$

velja:

$$\left|g(x,y,e^{-t}) - g(x,y,0)\right| = \left|\frac{\partial g}{\partial s}(x,y,s)\right| \leq \left(M_0 + M_2 + M_1\|x\|\right) e^{-\frac{1}{4}\|y-sx\|^2}$$

Sledi:

$$\begin{aligned}
\left|(P_t f)(x) - \int f\, d\gamma_p\right| &\leq e^{-t}(M_0 + M_2 + M_1\|x\|) \int_{\mathbb{R}^p} |f(y)| e^{-\frac{1}{4}\|y-sx\|^2} dy \leq \\
&\leq e^{-t}\|f\|_L (M_0 + M_2 + M_1\|x\|) \int_{\mathbb{R}^p} e^{c\|y\|} e^{-\frac{1}{4}\|y-sx\|^2}\, dy = \\
&= e^{-t}\|f\|_L (M_0 + M_2 + M_1\|x\|) \int_{\mathbb{R}^p} e^{c\|z+sx\|} e^{-\frac{1}{4}\|z\|^2} dz \leq \\
&\leq e^{-t}\|f\|_L (M_0 + M_2 + M_1\|x\|) e^{c\|x\|/e} \int_{\mathbb{R}^p} e^{\|z\| - \frac{1}{4}\|z\|^2} dz
\end{aligned}$$

Obstaja tak $M > 0$, da za vsak $s \geq 0$ velja:

$$(M_0 + M_2 + M_1 s) e^{-c(1-1/e)s} \leq M$$

Za vsak $t \geq 1$ in vsak $x \in \mathbb{R}^p$ torej velja:

$$e^{-c\|x\|}\left|(P_t f)(x) - \int f\, d\gamma_p\right| \leq e^{-t}\|f\|_L M \int_{\mathbb{R}^p} e^{\|z\| - \frac{1}{4}\|z\|^2} dz$$

oziroma:

$$\left\|P_t f - \int f\, d\gamma_p\right\|_L \leq e^{-t}\|f\|_L M \int_{\mathbb{R}^p} e^{\|z\| - \frac{1}{4}\|z\|^2} dz$$

Elementi $P_t f$ torej v prostoru L eksponentno konvergirajo proti $\int f\, d\gamma_p$, torej je gotovo $\int_0^\infty \|P_t f - \int f\, d\gamma_p\|_L\, dt < \infty$. Po izreku 1.0.1 potem za vsak $f \in L$ rešitev Steinove enačbe:

$$Gg = f - \int f\, d\mu$$

obstaja in se izraža v obliki:

$$g = -\int_0^\infty \left(P_t f - \int f\, d\gamma_p\right) dt$$

Opomba. Seveda se da Steinova enačba rešiti tudi bolj neposredno, brez uporabe izreka 1.0.1. Vse, kar je treba narediti, je, da rešitev, dobljeno z zgornjo formulo, vstavimo v Steinovo enačbo in preverimo, da drži. Pri tem bolj ali manj oponašamo dokaz izreka

1.0.1, le da moramo vse, kar v dokazu izreka 1.0.1 sledi iz teorije operatorskih polgrup, pokazati posebej: namesto na konvergenco v normi prostora L se skličemo na enakomerno konvergenco po kompaktih in tako izpeljemo ključno Dynkinovo formulo (trditev B.11.7) in dokažemo, da smemo odvajati pod integralskim znakom.

Nastane seveda vprašanje, ali je funkcija g v množici D, definirani v formuli (4.1.2). Prav tako se je smiselno vprašati, ali ima funkcija g omejene tretje parcialne odvode. Za vsoto neodvisnih in enako porazdeljenih slučajnih spremenljivk smo namreč v formuli (3.1.2) ocenili:

$$\left|\mathbf{E}\big(g''(X) - Xg'(X)\big)\right| \le n^{-1/2} M_3(g)\big(\tfrac{3}{2}\alpha + \tfrac{5}{6}\beta\big)$$

kjer po trditvi 3.1.1 velja $M_3(g) = \operatorname{ess\,sup}_x |g'''(x)|$. Podobna ocena se da narediti tudi v večrazsežnem primeru, le več dela je treba, glavna ideja pa je ista (glej [13]). Na obe vprašanji bomo odgovorili v naslednjem razdelku.

4.2 Ocenjevanje

Definicija. Funkcija $f\colon \mathbb{R}^p \to \mathbb{R}$ je *eksponentno omejena*, če obstaja tak $c \in \mathbb{R}$, da je $|f(x)| \le e^{c\|x\|}$ za vsak $x \in \mathbb{R}^p$.

Opomba. Očitno je vsaka funkcija eksponentno omejena natanko tedaj, ko pripada kakemu prostoru L, definiranem v prejšnjem razdelku (za primerno izbrano konstanto c).

Opomba. Če je $k \in \mathbb{N}_0$, $f \in \mathcal{C}^{(k)}(\mathbb{R}^p)$ in so vsi parcialni odvodi funkcije f reda k eksponentno omejeni, je tudi f eksponentno omejena.

Do konca razdelka naj bo f eksponentno omejena funkcija in naj bo še:

$$g := -\int_0^\infty \left(P_t f - \int f\, d\gamma_p\right) dt$$

Ker f pripada primernemu prostoru L, je ta integral vedno dobro definiran. Najprej bomo pokazali, da so parcialni odvodi funkcije g omejeni z istimi parcialnimi odvodi funkcije f. Kasneje bomo ta rezultat še izboljšali.

Trditev 4.2.1. *Naj bo $k \ge 1$ in naj bo $f \in \mathcal{C}^{(k)}(\mathbb{R}^p)$. Vsi njeni parcialni odvodi reda k naj bodo eksponentno omejeni. Tedaj je tudi $g \in \mathcal{C}^{(k)}(\mathbb{R}^p)$ in za poljuben parcialni odvod ∂_k reda k velja:*

$$\partial_k g = -\int_0^\infty e^{-kt} P_t \partial_k f\, dt \tag{4.2.1}$$

Opomba. Zgornji integral lahko gledamo bodisi kot integral po točkah bodisi kot integral v Banachovem prostoru L za primeren c. Seveda v obeh primerih dobimo isto (glej strani 114–117 in formulo (B.11.2)).

Dokaz trditve. Brez škode za splošnost lahko privzamemo, da je $\int f\, d\gamma_p = 0$. V tem primeru je naša trditev smiselno formulirana tudi za $k = 0$. Še več, za $k = 0$ je naša trditev tudi pravilna. Naredimo zdaj indukcijski korak s $k-1$ na k. Vsak parcialni odvod ∂_k reda k lahko zapišemo kot kompozitum $\partial\partial_{k-1}$, kjer je ∂ parcialni odvod prvega reda, ∂_{k-1} pa parcialni odvod reda $k-1$. Indukcijski korak naredimo tako, da odvajamo enačbo, ki drži po indukcijski predpostavki:

$$\partial_{k-1} g = -\int_0^\infty e^{-(k-1)t} P_t \partial_{k-1} f \, dt$$

Nastane vprašanje, ali smemo odvajati pod integralskim znakom. Po lemi B.13.2 velja:

$$\partial e^{-(k-1)t} P_t \partial_{k-1} f = e^{-kt} P_t \partial_k f$$

Integral:

$$-\int_0^\infty e^{-kt} (P_t \partial_k f)(x)$$

pa konvergira enakomerno na vseh kompaktnih območjih, saj je funkcija $\partial_k f$, z njo pa tudi $P_t \partial_k f$, eksponentno omejena. Zato lahko odvajamo pod integralskim znakom, torej je zgornji integral res enak $\partial_k g$. Integral pa lahko zapišemo tudi v obliki:

$$-\frac{1}{(2\pi)^{p/2}} \int_0^\infty \int_{\mathbb{R}^p} e^{-kt} (\partial_k f) \left(e^{-t} x + \sqrt{1 - e^{-2t}}\, y \right) e^{-\frac{1}{2}\|y\|^2} dy \, dt$$

kar je $(p+1)$-kratni integral zvezne funkcije spremenljivk x, y in t. Ta integral glede na x konvergira enakomerno po vseh kompaktnih območjih, zato je tudi sam zvezna funkcija spremenljivke x. To pa je vse, kar je bilo treba dokazati. ■

Posledica 4.2.2. *Brž ko je $f \in \mathcal{C}^{(2)}(\mathbb{R}^p)$ in so drugi parcialni odvodi funkcije f eksponentno omejeni, funkcija g pripada množici D, definirani v formuli (4.1.2) (za primerno izbrano konstanto c v definiciji prostora L), torej je g rešitev enačbe:*

$$(\Delta g)(x) - x \cdot (\nabla g)(x) = f(x) - \int f \, d\gamma_p$$

Če namreč integral v formuli (4.2.1) gledamo kot integral v Banachovem prostoru L, dobimo, da v primeru, ko funkcija f s svojimi prvimi in drugimi parcialnimi odvodi vred pripada prostoru L za primerno izbrano konstanto c, to velja tudi za g. Za vsak večji c zdaj dobimo $f \in D$.

Posledica 4.2.3. *Če so vsi parcialni odvodi reda k omejeni, velja ocena:*

$$\|\partial_k g\|_\infty \leq \frac{1}{k} \|\partial_k f\|_\infty$$

Če torej za funkcijo $h \in \mathcal{C}^{(k)}(\mathbb{R}^p)$ označimo:

$$M_k(h) := \max_{\partial_k} \|\partial_k h\|_\infty$$

kjer ∂_k preteče vse parcialne odvode reda k, velja ocena:

$$M_k(g) \leq \frac{M_k(f)}{k}$$

Iz teh ocen takoj sledi naslednja trditev.

Trditev 4.2.4. *Naj bo X slučajna spremenljivka in naj za vsako funkcijo $g \in \mathcal{C}^{(3)}(\mathbb{R}^p)$ z omejenimi tretjimi parcialnimi odvodi velja:*

$$\left|\mathbf{E}\big((\Delta g)(X) - X \cdot (\nabla g)(X)\big)\right| \le c\, M_3(g) \tag{4.2.2}$$

(glej konec prejšnjega razdelka). *Tedaj za vsako funkcijo $f \in \mathcal{C}^{(3)}(\mathbb{R}^p)$ z omejenimi tretjimi parcialnimi odvodi velja:*

$$\left|\int f\, d\mathcal{L}(X) - \int f\, d\gamma_p\right| = \left|\mathbf{E}\big(f(X)\big) - \int f\, d\gamma_p\right| \le \tfrac{1}{3} c\, M_3(f)$$

Ta ocena nam že da dobro merilo, kako blizu je porazdelitev slučajne spremenljivke X normalni porazdelitvi. Seveda pa nam ta ocena sama po sebi kaj malo pomaga, če želimo dobiti npr. oceno v Wassersteinovi metriki. Prvi korak k temu bo izboljšava posledice 4.2.3. Pokazali bomo, da v primeru, ko je $f \in \mathcal{C}^k(\mathbb{R}^p)$ in so vsi parcialni odvodi funkcije f reda k omejeni, velja, da je $g \in \mathcal{C}^{(k+1)}(\mathbb{R}^p)$ in da so omejeni vsi parcialni odvodi funkcije g reda $k+1$. To sicer še ne bo dovolj za oceno v Wassersteinovi metriki, saj bomo dobili oceno z $M_2(f)$, želeli pa bi oceno z $M_1(f)$. Vendar pa se bo dalo popraviti tudi to, in sicer tako, da bomo funkcijo f še pred uporabo Chen–Steinove metode malo zgladili. Natančneje, preslikali jo bomo z elementi T_t operatorske polgrupe Brownovega gibanja, torej:

$$(T_t f)(x) = \int f\big(x + \sqrt{t}\, y\big)\, \gamma_p(dy)$$

Lema 4.2.5. *Če je f omejena merljiva funkcija in $t > 0$, je $T_t f \in \mathcal{C}^{(1)}(\mathbb{R}^p)$ in velja:*

$$\left(\frac{\partial}{\partial x_i} T_t f\right)(x) = \frac{1}{\sqrt{t}} \int f\big(x + \sqrt{t}\, y\big) y_i\, \gamma_p(dy)$$

Dokaz. Najprej velja:

$$(T_t f)(x) = \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} f\big(x + \sqrt{t}\, y\big) e^{-\frac{1}{2}\|y\|^2} dy = \frac{1}{(2\pi t)^{p/2}} \int_{\mathbb{R}^p} f(z)\, e^{-\frac{1}{2}\frac{\|z-x\|^2}{t}} dz$$

Nastane vprašanje, ali lahko to odvajamo pod integralskim znakom. Velja:

$$\frac{\partial}{\partial x_i}\left(e^{-\frac{1}{2}\frac{\|z-x\|^2}{t}}\right) = \frac{z_i - x_i}{t} e^{-\frac{1}{2}\frac{\|z-x\|^2}{t}}$$

Funkcija f je omejena. Če je K kompaktna množica v $\mathbb{R}^p$, velja:

$$\int_{\mathbb{R}^p} \sup_{x \in K} \left|\frac{z_i - x_i}{t} e^{-\frac{1}{2}\frac{\|z-x\|^2}{t}}\right| dz < \infty$$

kar ni težko preveriti. To pa pomeni, da lahko pri odvajanju integrala uporabimo izrek o dominirani konvergenci. Sledi:

$$\left(\frac{\partial}{\partial x_i}T_t f\right)(x) = \frac{1}{(2\pi)^{p/2}t^{1+p/2}}\int_{\mathbb{R}^p} f(z)(z_i - x_i)e^{-\frac{1}{2}\frac{\|z-x\|^2}{t}}dz$$

S ponovno uporabo izreka o dominirani konvergenci se prepričamo, da je zgornja funkcija zvezna. Po substituciji dobimo:

$$\left(\frac{\partial}{\partial x_i}T_t f\right)(x) = \frac{1}{(2\pi)^{p/2}\sqrt{t}}\int_{\mathbb{R}^p} f\big(x+\sqrt{t}\,y\big)y_i e^{-\frac{1}{2}\|y\|^2}dy = \frac{1}{\sqrt{t}}\int f\big(x+\sqrt{t}\,y\big)y_i\,\gamma_p(dy)$$

To pa je bilo treba dokazati. ■

Opomba. V resnici velja celo $T_t f \in \mathcal{C}^\infty(\mathbb{R}^p)$. To vidimo tako, da element T_t zapišemo kot potenco $T_{t/n}^n$. Iz prejšnje leme in leme B.12.6 se da izpeljati, da operatorji $T_{t/n}$ slikajo prostor $\mathcal{C}^{(k)}(\mathbb{R}^p)$ v prostor $\mathcal{C}^{(k+1)}(\mathbb{R}^p)$, torej mora $T_t f$ pripadati $\mathcal{C}^{(n)}(\mathbb{R}^p)$ za vsak $n \in \mathbb{N}$.

Posledica 4.2.6. *Če je f merljiva in omejena, velja:*

$$\left\|\frac{\partial}{\partial x_i}T_t f\right\|_\infty \le \frac{1}{\sqrt{t}}\|f\|_\infty\int|y_i|\,\gamma_p(dy) = \frac{1}{\sqrt{t}}\|f\|_\infty\int|s|\,\gamma_1(ds) = \sqrt{\frac{2}{\pi}}\,\frac{1}{\sqrt{t}}\,\|f\|_\infty$$

kjer je $\|f\|_\infty = \operatorname{ess\,sup}_{x\in\mathbb{R}^p}|f(x)|$.

Izboljšajmo zdaj posledico 4.2.3 tako, da bomo $M_{k+1}(g)$ ocenili z $M_k(f)$. Naj bo torej $k \ge 1$, $f \in \mathcal{C}^{(k)}(\mathbb{R}^p)$ in $M_k(f) < \infty$. Naj bo ∂_{k+1} kak parcialni odvod reda $k+1$. Zapišemo lahko $\partial_{k+1} = \partial\partial_k$, kjer je ∂_k parcialni odvod reda k, ∂ pa reda 1. Po trditvi 4.2.1 velja:

$$\partial_k g = -\int_0^\infty e^{-kt}P_t\partial_k f\,dt \tag{4.2.3}$$

Odvajajmo zdaj to enačbo parcialno po x_i. Spet nastane vprašanje, ali smemo odvajati pod integralskim znakom. Če želimo dobiti odgovor na to, moramo seveda odvajati integrand. Če označimo $(v_t f)(x) := f(tx)$, velja $P_t = v_{e^{-t}}T_{1-e^{-2t}}$. Torej velja:

$$\partial P_t\partial_k f = \partial v_{e^{-t}}T_{1-e^{-2t}} = e^{-t}v_{e^{-t}}\partial T_{1-e^{-2t}} \tag{4.2.4}$$

Iz posledice 4.2.6 sledi:

$$\left\|e^{-kt}\partial P_t\partial_k f\right\|_\infty = \left\|e^{-(k+1)t}\partial T_{1-e^{-2t}}\partial_k f\right\|_\infty \le \sqrt{\frac{2}{\pi}}\,\frac{e^{-(k+1)t}}{\sqrt{1-e^{-2t}}}\|\partial_k f\|_\infty \tag{4.2.5}$$

Ker je:

$$\int_0^\infty \frac{e^{-(k+1)t}}{\sqrt{1-e^{-2t}}}\,dt = \int_0^1 \frac{u^k}{\sqrt{1-u^2}}\,du = \frac{1}{2}\int_0^1 s^{(k-1)/2}(1-s)^{-1/2}ds = \frac{\Gamma\big(\frac{k+1}{2}\big)\sqrt{\pi}}{2\Gamma\big(\frac{k}{2}+1\big)} < \infty$$

integral odvoda integranda v formuli (4.2.3) enakomerno konvergira. Od tod pa sledi, da lahko odvajamo pod integralskim znakom. Velja torej:

$$\partial_{k+1}g = -\int_0^\infty \frac{\partial}{\partial x_i}\big(e^{-kt}P_t\partial_k f\big)\,dt$$

Iz (4.2.4) in trditve 4.2.5 sledi, da je integrand zvezna funkcija. Ker integral enakomerno konvergira, je zvezna tudi funkcija $\partial_{k+1} g$. Iz ocene (4.2.5) pa sledi:

$$\|\partial_{k+1} g\|_\infty \le \sqrt{\frac{2}{\pi}}\,\|\partial_k f\|_\infty \int_0^\infty \frac{e^{-(k+1)t}}{\sqrt{1-e^{-2t}}}\,dt = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{k}{2}+1\right)}\|\partial_k f\|_\infty$$

Dokazali smo naslednjo trditev.

Trditev 4.2.7. *Naj bo $k \ge 1$ in naj bo $f \in \mathcal{C}^{(k)}(\mathbb{R}^p)$. Naj bo $M_k(f) < \infty$. Tedaj je $g \in \mathcal{C}^{(k+1)}(\mathbb{R}^p)$ in velja ocena:*

$$M_{k+1}(g) \le \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{k}{2}+1\right)}\,M_k(f)$$

Opomba. V zgornji trditvi odvode funkcije g res ocenjujemo z enega reda nižjimi odvodi funkcije f, je pa ta ocena asimptotično slabša. Je namreč reda $k^{-1/2}$, medtem ko je ocena iz posledice 4.2.3 velikostnega reda k^{-1}.

Opomba. Človek bi se vprašal, ali se da z $M_k(f)$ oceniti tudi $M_{k+2}(g)$, saj smo pri enorazsežni normalni aproksimaciji z $M_1(f)$ ocenili $M_3(g)$ (glej trditev 3.2.2). Konec koncev opomba k trditvi 4.2.5 pove, da funkcija $T_t f$ ne pripada le $\mathcal{C}^{(1)}(\mathbb{R}^p)$, temveč tudi $\mathcal{C}^{(2)}(\mathbb{R}^p)$. In v resnici se da z $\|f\|_\infty$ oceniti tudi druge odvode funkcije $T_t f$, le da je tokrat ocena velikostnega reda $\frac{1}{t}$. Žal je to usodno za nadaljnje ocenjevanje, saj dobimo divergenten integral. Vprašanje, ali se da $M_{k+2}(g)$ oceniti že z $M_k(f)$, puščamo odprto.

Za $k = 2$ dobimo:

$$M_3(g) \le \sqrt{\frac{\pi}{8}}\,M_2(f)$$

in tako dobimo naslednjo izboljšavo trditve 4.2.4.

Trditev 4.2.8. *Naj bo X slučajna spremenljivka in naj za vsako funkcijo $g \in \mathcal{C}^{(3)}(\mathbb{R}^p)$ z omejenimi tretjimi parcialnimi odvodi velja:*

$$\left|\mathbf{E}\big((\Delta g)(X) - X\cdot(\nabla g)(X)\big)\right| \le c\,M_3(g) \tag{4.2.6}$$

Tedaj za vsako funkcijo $f \in \mathcal{C}^{(2)}(\mathbb{R}^p)$ z omejenimi drugimi parcialnimi odvodi velja:

$$\left|\int f\,d\mathcal{L}(X) - \int f\,d\gamma_p\right| = \left|\mathbf{E}\big(f(X)\big) - \int f\,d\gamma_p\right| \le \sqrt{\frac{\pi}{8}}\,c\,M_2(f)$$

Seveda pa to še ni dovolj za ocenjevanje v Wassersteinovi metriki. Vendar pa je do tega zdaj le še majhen korak. Kot smo že omenili, moramo pred uporabo Chen–Steinove metode funkcijo f najprej malo zgladiti. Naj bo f Lipschitzova s konstanto 1. Tedaj velja:

$$(T_t f)(x) - f(x) = \int \Big(f\big(x + \sqrt{t}\,y\big) - f(x)\Big)\gamma_p(dy)$$

Sledi:

$$\|T_t f - f\|_\infty \le \sqrt{t} \int \|y\| \, \gamma_p(dy)$$

Velja:

$$\int \|y\| \, \gamma_p(dy) = \mathbf{E}\big(\sqrt{Z}\big)$$

kjer ima Z porazdelitev hi kvadrat s p prostostnimi stopnjami, le-ta pa ima gostoto:

$$p_Z(z) = \frac{1}{2^{p/2}\Gamma\big(\frac{p}{2}\big)} t^{p/2-1} e^{-t/2}$$

za $z > 0$. Sledi:

$$\int \|y\| \, \gamma_p(dy) = \int_0^\infty \sqrt{z} \, p_Z(z) \, dz = \sqrt{2} \, \frac{\Gamma\big(\frac{p+1}{2}\big)}{\Gamma\big(\frac{p}{2}\big)} \approx \sqrt{2p}$$

torej je:

$$\|T_t f - f\|_\infty \le \sqrt{2t} \, \frac{\Gamma\big(\frac{p+1}{2}\big)}{\Gamma\big(\frac{p}{2}\big)} \tag{4.2.7}$$

Oglejmo si, kaj smo z glajenjem pridobili! Naj bo $1 \le i, j \le p$. Ker je f Lipschitzova s konstanto 1, je $\frac{\partial f}{\partial x_j}$ merljiva funkcija, ki po absolutni vrednosti skoraj nikjer ne preseže 1. Podobno kot v lemi B.12.6 lahko tudi tu dokažemo, da velja:

$$\frac{\partial}{\partial x_j} T_t f = T_t \frac{\partial}{\partial x_j} f$$

povsod tam, kjer je f odvedljiva po x_j. Ključno vprašanje je spet odvajanje pod integralskim znakom in iz dejstva, da je f Lipschitzova s konstanto 1, sledi, da lahko uporabimo izrek o dominirani konvergenci (v dokazu leme B.12.6 pa to utemeljimo z Lagrangeovim izrekom). Iz posledice 4.2.6 zdaj sledi:

$$\left\| \frac{\partial^2 (T_t f)}{\partial x_i \, \partial x_j} \right\|_\infty = \left\| \frac{\partial}{\partial x_i} \left(T_t \frac{\partial}{\partial x_j} \right) \right\|_\infty \le \sqrt{\frac{2}{\pi t}}$$

torej je:

$$M_2(T_t f) \le \sqrt{\frac{2}{\pi t}}$$

Naj zdaj velja (4.2.6). Iz zgornje ocene, trditve 4.2.8 in ocene (4.2.7) zdaj dobimo:

$$\begin{aligned}
\left| \int f \, d\big(\mathcal{L}(X) - \gamma_p\big) \right| &\le \left| \int (T_t f - f) \, d\big(\mathcal{L}(X) - \gamma_p\big) \right| + \left| \int T_t f d\big(\mathcal{L}(X) - \gamma_p\big) \right| \le \\
&\le 2\|T_t f - f\| + \left| \int T_t f d\big(\mathcal{L}(X) - \gamma_p\big) \right| \le \\
&\le 2\sqrt{2} \, \frac{\Gamma\big(\frac{p+1}{2}\big)}{\Gamma\big(\frac{p}{2}\big)} \sqrt{t} + \sqrt{\frac{\pi}{8}} \, c \, M_2(T_t f) \le \\
&\le 2\sqrt{2} \, \frac{\Gamma\big(\frac{p+1}{2}\big)}{\Gamma\big(\frac{p}{2}\big)} t^{1/2} + \frac{1}{2} c \, t^{-1/2}
\end{aligned}$$

Poiskati moramo le še najugodnejši t. Funkcija:

$$h(t) := at^{1/2} + bt^{-1/2} = (\sqrt{a}\,t^{1/4} - \sqrt{b}\,t^{-1/4})^2 + 2\sqrt{ab}$$

ima najmanjšo vrednost $2\sqrt{ab}$, torej končno velja:

$$\left|\int f\,d\mathcal{L}(X) - \int f\,d\gamma_p\right| \le \sqrt{2\,\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}\,c}$$

Dokazali smo naslednjo trditev, ki nam da oceno v Wassersteinovi metriki.

Trditev 4.2.9. *Naj bo X slučajna spremenljivka in naj za vsako funkcijo $g \in \mathcal{C}^{(3)}(\mathbb{R}^p)$ z omejenimi tretjimi parcialnimi odvodi velja:*

$$\left|\mathbf{E}\big((\Delta g)(X) - X \cdot (\nabla g)(X)\big)\right| \le c\,M_3(g) \tag{4.2.8}$$

Tedaj velja:

$$d_{\mathrm{W}}\big(\mathcal{L}(X), \gamma_p\big) \le \sqrt{2\,\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}\,c} \approx \sqrt{2c}\,p^{1/4}$$

Zgornjo trditev smo navedli zgolj za ilustracijo ocenjevanja Wassersteinove metrike prek Chen–Steinove metode. Ta način ocenjevanja deluje tudi, če imamo namesto ocene (4.2.8) na razpolago naslednjo oceno:

$$\left|\mathbf{E}\big((\Delta g)(X) - X \cdot (\nabla g)(X)\big)\right| \le c_1 M_1(g) + c_2 M_2(g) + c_3 M_3(g)$$

Pripomniti pa velja, da v resnici ocene s količinami M_k, kot smo jih definirali (torej kar s supremumi parcialnih odvodov) niso najbolj naravne. Naravneje bi bilo k-ti odvod dane funkcije h, ki ga označimo z $D^k h$, gledati kot simetrično k-linearno formo in definirati:

$$M_k(h) := \sup_{\substack{x \in \mathbb{R}^p \\ \|v_i\|=1 \\ i=1,\dots k}} \left|(D^k h)(x)(v_1, \dots v_k)\right|$$

Tudi pri tako definiranih količinah M_k bi lahko izpeljali trditve, podobne trditvam 4.2.4, 4.2.8 in 4.2.9.

Pri izpeljevanju smo šli do ocene v Wassersteinovi metriki. Kot smo že omenili, pa je v članku [13] izpeljan tudi izrek Berry–Esseenovega tipa: testne funkcije so indikatorji merljivih konveksnih množic. Še več, izpeljane so ocene za precej splošne razrede testnih funkcij. Osnovna ideja izpeljave je sicer ista, izvedba pa je dosti težja kot pri oceni v Wassersteinovi metriki.

Dodatek A

Konvergenca porazdelitev

Dostikrat nas zanima, ali zaporedje porazdelitev $(\mu_n)_{n\in\mathbb{N}}$ na danem merljivem prostoru $(S, \mathcal{S})$ konvergira k dani porazdelitvi μ. Za ta namen pa moramo na prostoru $\mathrm{Pr}(S, \mathcal{S})$ vseh porazdelitev na $(S, \mathcal{S})$ najprej definirati topologijo. To pa storimo tako, da porazdelitve "otipavamo" s primernimi *testnimi funkcijami*: porazdelitvi μ in ν sta si "blizu", če sta blizu integrala $\int f\,d\mu$ in $\int f\,d\nu$ za primeren nabor testnih funkcij f.

Topologijo lahko definiramo na dva načina. Lahko definiramo kar metriko:

$$d(\mu, \nu) = \sup_{f\in\mathcal{T}} \left| \int f\,d\nu - \int f\,d\mu \right| \tag{A.0.1}$$

kjer je $\mathcal{T}$ naš razred testnih funkcij. Nastane vprašanje, ali je to res metrika. Simetrija je očitna, tudi trikotniško neenakost preverimo brez težav. Zatakne pa se pri preverjanju, ali iz $d(\mu, \nu) = 0$ sledi $\mu = \nu$. Odgovor ne bo vedno pozitiven. Če npr. za $\mathcal{T}$ vzamemo vse konstantne funkcije, je $d(\mu, \nu) = 0$ za poljubni porazdelitvi μ in ν. Da bo d res metrika, mora biti razred $\mathcal{T}$ dovolj bogat.

Trditev A.0.1. *Naj bo $\mathcal{P}$ π-sistem, ki generira $\mathcal{S}$, in naj za vsak $A \in \mathcal{P}$ obstaja enakomerno omejeno zaporedje funkcij $f_n \in \mathcal{T}$, ki po točkah konvergira k indikatorju I_A. Tedaj je d res metrika.*

Dokaz. Naj bo $d(\mu, \nu) = 0$, torej je $\int f\,d\mu = \int f\,d\nu$ za vsak $f \in \mathcal{T}$. Iz izreka o dominirani konvergenci takoj dobimo, da je $\mu(A) = \nu(A)$ za vsak $A \in \mathcal{P}$. Družina vseh množic A, za katere je $\mu(A) = \nu(A)$, pa je λ-sistem. Pravkar smo dokazali, da ta λ-sistem vsebuje π-sistem $\mathcal{P}$, ki generira $\mathcal{S}$. Po Dynkinovi lemi mora biti potem $\mu(A) = \nu(A)$ za vsak $A \in \mathcal{S}$, torej je res $\mu = \nu$. ■

Druga možnost pa je, da vpeljemo topologijo s podbazo, ki jo sestavljajo množice oblike:

$$G(f, U) := \left\{ \nu \mid \textstyle\int f\,d\nu \in U \right\} \tag{A.0.2}$$

kjer je $f \in \mathcal{T}$, U pa je odprta množica v $\mathbb{R}$. Brez težav preverimo, da v tej topologiji za vsako porazdelitev μ množice oblike:

$$N(\mu; f_1, \dots f_n; \varepsilon_1, \dots \varepsilon_n) := \left\{ \nu \mid \left| \textstyle\int f_i\,d\nu - \int f_i\,d\mu \right| < \varepsilon_i, i = 1, \dots n \right\} \tag{A.0.3}$$

kjer je $f_1, \dots f_n \in t$ in $\varepsilon_1, \dots \varepsilon_n > 0$, sestavljajo bazični sistem okolic porazdelitve μ. *Od tod dobimo, da v tako definirani topologiji zaporedje porazdelitev μ_n konvergira proti* *porazdelitvi μ natanko tedaj, ko za vsako funkcijo $f \in \mathcal{T}$ zaporedje $\int f\, d\mu_n$ konvergira proti $\int f\, d\mu$.*

Za razred testnih funkcij $\mathcal{T}$ smo torej topologijo vpeljali na dva načina. Seveda pa ni nikjer rečeno, da sta ekvivalentna. Velja pa vsebovanost v eno smer.

Trditev A.0.2. *Topologija, vpeljana z metriko iz formule (A.0.1), je močnejša od topologije, definirano s podbazo iz formule (A.0.2).*

DOKAZ. Označimo s $K(\mu, r)$ odprto kroglo okrog porazdelitve μ s polmerom r. Velja torej $K(\mu, r) = \{\nu \mid d(\mu, \nu) < r\}$. Tedaj je dovolj dokazati, da za vsako množico $N(\mu; f_1, \dots f_n; \varepsilon_1, \dots \varepsilon_n)$ obstaja odprta krogla oblike $K(\mu, r)$, ki je vsebovana v njej. To pa je gotovo res, če postavimo $r := \min\{\varepsilon_1, \dots \varepsilon_n\}$. ∎

Posledica A.0.3. *Če zaporedje porazdelitev $(\mu_n)_{n\in\mathbb{N}}$ konvergira proti μ bodisi v topologiji, ki jo porodi metrika iz formule (A.0.1), bodisi v topologiji, definirani s podbazo iz formule (A.0.2), za vsak $f \in \mathcal{T}$ velja $\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$.*

Pripomnimo naj še, da tako definirane metrike in topologije še zdaleč niso nujno definirane na vseh merljivih prostorih. Navadno delamo v prostorih, ki so opremljeni vsaj še s kako lepo topologijo, če ne že z metriko. In tudi na določenem prostoru ni nujno, da definiramo topologijo na prostoru *vseh* porazdelitev. Dostikrat se moramo omejiti le na določen razred.

A.1 Metrika totalne variacije

Razred testnih funkcij, ki ga lahko definiramo v vsakem primeru, so seveda indikatorji. Le-ta nam določa *metriko totalne variacije*. Za dani porazdelitvi μ in ν torej definiramo:

$$d_{\mathrm{TV}}(\mu, \nu) := \sup_{A \in \mathcal{S}} |\nu(A) - \mu(A)|$$

Brez težav preverimo, da je to res metrika. Definirana je na prostoru porazdelitev vsakega merljivega prostora. Vendar pa se izkaže, da je ta metrika navadno daleč premočna. Oglejmo si recimo primer, ko je S topološki prostor z Borelovo σ-algebro in zaporedje točk x_n, ki niso enake x, konvergira k x. Kljub temu pa zaporedje Diracovih mer δ_{x_n} ne konvergira k δ_x, torej niti skoraj gotova konvergenca (ki je zelo močna) ne implicira konvergence v tej metriki. Metrika totalne variacije je zato dobra le v primeru, ko je prostor S diskreten.

Pri metriki totalne variacije lahko za testne funkcije vzamemo tudi vse merljive funkcije iz S v $[0, 1]$. Velja namreč:

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{0 \le f \le 1} \left| \int f\, d\nu - \int f\, d\mu \right| \tag{A.1.1}$$

Da je levi izraz manjši ali enak desnemu, je očitno, saj razred $\mathcal{T}$ vsebuje vse indikatorje. V obratno smer pa dokažemo tako, da definiramo še en, vmesni razred testnih funkcij. To bo razred $\mathcal{E}$ merljivih funkcij f, za katere obstaja tak $n \in \mathbb{N}$, da f zavzame le vrednosti

$\frac{k}{n}$, $k = 0, \ldots n$. Naj bosta μ in ν porazdelitvi in naj bo $r := d_{\mathrm{TV}}(\mu, \nu)$. Vsako funkcijo iz $\mathcal{E}$ lahko zapišemo v obliki:

$$f = \frac{1}{n} \sum_{k=1}^{n} I_{A_k}$$

Sledi:

$$\left| \int f\, d\nu - \int f\, d\mu \right| \leq r \tag{A.1.2}$$

Končno, naj bo f poljubna merljiva funkcija iz S v $[0, 1]$. Obstaja zaporedje f_n funkcij iz $\mathcal{E}$, ki po točkah konvergira k f. Po izreku o dominirani konvergenci potem neenačba (A.1.2) velja za vse funkcije iz $\mathcal{T}$. Od tod pa že sledi enačba (A.1.1).

Opomba. Nekateri avtorji za metriko totalne variacije jemljejo tudi naslednjo metriko:

$$d(\mu, \nu) = \sup_{\|f\| \leq 1} \left| \int f\, d\nu - \int f\, d\mu \right| = 2 d_{\mathrm{TV}}(\mu, \nu)$$

A.2 Šibka topologija

V primeru, ko je prostor S opremljen s topologijo in je $\mathcal{S}$ Borelova σ-algebra, lahko prostor $\mathrm{Pr}(S, \mathcal{S})$ opremimo s *šibko topologijo*, pri kateri za razred testnih funkcij vzamemo $\mathcal{C}_\flat(S)$, t. j. razred vseh zveznih omejenih funkcij na S. Tu pa topologijo definiramo s podbazo iz formule (A.0.2). V tej topologiji torej zaporedje $(\mu_n)_{n \in \mathbb{N}}$ konvergira k μ natanko tedaj, ko za vsak $f \in \mathcal{C}_\flat(S)$ zaporedje $\int f\, d\mu_n$ konvergira k $\int f\, d\mu$. Rekli bomo, da zaporedje porazdelitev μ_n *šibko konvergira* k μ.

Opomba. *Šibka topologija je šibkejša od tiste, ki jo porodi metrika totalne variacije.* Gotovo je namreč šibkejša od topologije, definirane s podbazo iz formule (A.0.2), pri čemer za razred testnih funkcij vzamemo vse omejene merljive funkcije. Isto topologijo dobimo, če vzamemo le vse merljive funkcije iz S v $[0, 1]$, le-ta pa je po trditvi A.0.2 šibkejša od topologije, ki jo porodi metrika totalne variacije.

Opomba. Obratno seveda ni res, saj zaporedje Diracovih mer δ_{x_n} vedno šibko konvergira proti x_n, če x_n konvergira proti x. V metriki totalne variance pa to zaporedje konvergira le, če za dovolj velike n velja $x_n = x$.

Konvergenca v šibki topologiji se da karakterizirati z odprtimi in zaprtimi množicami. Velja namreč naslednji izrek.

Izrek A.2.1. *Naj bodo μ in μ_n, $n \in \mathbb{N}$, porazdelitve na topološkem prostoru S z Borelovo σ-algebro. Naslednje trditve so ekvivalentne:*

(1) *Zaporedje $(\mu_n)_{n \in \mathbb{N}}$ šibko konvergira k μ.*

(2) *Za vsako zaprto množico $F \subset S$ velja $\limsup_{n \in \mathbb{N}} \mu_n(F) \leq \mu(F)$.*

(3) *Za vsako odprto množico $G \subset S$ velja $\liminf_{n \in \mathbb{N}} \mu_n(G) \geq \mu(F)$.*

Skica dokaza. Ekvivalenca (2) $\iff$ (3) je očitna. Naj velja (1) in naj bo F zaprta množica. Definirajmo funkcije $f_n\colon S \to \mathbb{R}$ po predpisu:

$$f_n(x) := \max\{0, 1 - n\, d(x, F)\}$$

Očitno je indikator I_F limita padajočega zaporedja funkcij f_n, ki so zvezne in omejene (še več, so celo Lipschitzove). Za vsak n velja:

$$\limsup_{m\to\infty} \mu_m(F) \le \limsup_{m\to\infty} f_n\, d\mu_m = \int f_n\, d\mu$$

Če pošljemo še n proti neskončno, po izreku o dominirani konvergenci dobimo, da je tudi $\limsup_{m\to\infty} \mu_m(F) \le \mu(F)$, torej velja (2).

Dokaz obratne implikacije je nekoliko težji in ga bomo opustili. Ideja je v tem, da zvezne funkcije aproksimiramo z enostavnimi (stopničastimi). Podrobnosti si lahko bralec pogleda npr. v [21] na strani 208, izrek 83.4. ■

Posledica A.2.2. *Če zaporedje $(\mu_n)_{n\in\mathbb{N}}$ šibko konvergira proti μ in je A taka Borelova množica v S, da je $\mu(\overline{A} \setminus \operatorname{Int} A) = 0$, velja $\lim_{n\to\infty} \mu_n(A) = 0$.*

Trditev A.2.3. *Naj bo S regularen topološki prostor s števno bazo (po Urisonovem metrizacijskem izreku je tak prostor metrizabilen) in Borelovo σ-algebro $\mathcal{S}$. Tedaj je šibka topologija na $\Pr(S, \mathcal{S})$ Hausdorffova.*

Dokaz. Najprej dokažimo tole trditev: če sta μ in ν porazdelitvi na $(S, \mathcal{S})$ in za vsako funkcijo $f \in \mathcal{C}_\flat(S)$ velja $\int f\, d\mu = \int f\, d\nu$, potem je $\mu = \nu$. Po trditvi A.0.1 je dovolj dokazati, da obstaja tak π-sistem $\mathcal{P}$, da za vsak $A \in \mathcal{P}$ obstaja enakomerno omejeno zaporedje zveznih funkcij, ki konvergirajo k indikatorju I_A. To pa ugotovimo tako, da za π-sistem vzamemo družino vseh zaprtih množic, za funkcije pa primerne Urisonove funkcije.

Naj bo zdaj $\mu \ne \nu$. Po prejšnjem obstaja taka funkcija $f \in \mathcal{C}_\flat(S)$, da je $a := \int f\, d\mu \ne \int f\, d\nu =: b$. Obstaja tak $\varepsilon > 0$, da je $(a - \varepsilon, a + \varepsilon) \cap (b - \varepsilon, b + \varepsilon) = \emptyset$. Potem pa je (s prejšnjimi oznakami) tudi $N(\mu, f, \varepsilon) \cap N(\nu, f, \varepsilon) = \emptyset$. ■

Hausdorffova lastnost je vse, kar bomo potrebovali od lastnosti šibke topologije na prostoru porazdelitev. Kot zanimivost pa naj omenimo, da velja še precej več.

Izrek A.2.4. *Naj bo S separabilen metrični prostor z Borelovo σ-algebro $\mathcal{S}$. Tedaj je prostor $\Pr(S, \mathcal{S})$ s šibko topologijo metrizabilen.*

Dokaz bomo opustili. Za primer, ko je prostor S *poln*, si ga lahko bralec pogleda v [21] na straneh 205–209. Ta dokaz je zelo posreden (skliče se npr. na Banach–Alaoglujev izrek). Vendar pa se da izrek dokazati tudi neposredno, in sicer tako, da prostor $\Pr(S, \mathcal{S})$ opremimo z *metriko Prohorova*:

$$\rho(\mu, \nu) := \inf\{\varepsilon \ge 0 \mid \mu(A) \le \nu(A^\varepsilon) + \varepsilon \text{ za vsako zaprto množico } A\}$$

kjer je $A^\varepsilon := \{x \in S \mid d(x, A) < \varepsilon\}$, d pa je metrika na S. Po precej mučnega računanja najprej preverimo, da je to res metrika in da res porodi šibko topologijo. Dobršen del tega si lahko bralec pogleda v [10] na straneh 96–110. Tam je dokazano vse razen dejstva, da je šibka topologija močnejša od topologije, ki jo porodi metrika Prohorova. To je treba preveriti neposredno, brez uporabe zaporedij.

Iz izreka o dominirani konvergenci takoj sledi naslednja trditev.

Trditev A.2.5. *Naj zaporedje slučajnih spremenljivk $(X_n)_{n\in\mathbb{N}}$ z vrednostmi v topološkem prostoru z Borelovo σ-algebro skoraj gotovo konvergira proti slučajni spremenljivki X. Tedaj zaporedje njihovih porazdelitev $\mathcal{L}(X_n)$ šibko konvergira k $\mathcal{L}(X)$.* ∎

Posledica A.2.6. *Naj zaporedje $(a_n)_{n\in\mathbb{N}}$ vektorjev v $\mathbb{R}^n$ konvergira proti A in naj zaporedje pozitivno semidefinitnih matrik $(\Sigma_n)_{n\in\mathbb{N}}$ iz $\mathbb{R}^{n\times n}$ konvergira proti Σ. Tedaj zaporedje normalnih porazdelitev $\mathrm{N}(a_n, \Sigma_n)$ šibko konvergira proti porazdelitvi $\mathrm{N}(a, \Sigma)$.*

Dokaz. Naj bo Z standardizirano normalno porazdeljena slučajna spremenljivka z vrednostmi v $\mathbb{R}^n$. Definirajmo slučajne spremenljivke X_n po predpisu $X_n := a_n + \Sigma_n^{1/2} Z$ in naj bo še $X := a + \Sigma^{1/2} Z$. Očitno je $X_n \sim \mathrm{N}(a_n, \Sigma_n)$ in $X \sim \mathrm{N}(a, \Sigma)$. Ker je tudi $\lim_{n\to\infty} \Sigma_n^{1/2} = \Sigma^{1/2}$, zaporedje slučajnih spremenljivk X_n po točkah konvergira k X, torej zaporedje njihovih porazdelitev šibko konvergira k $\mathrm{N}(a, \Sigma)$. ∎

A.3 Metrika Kolmogorova

Porazdelitve na realni osi lahko karakteriziramo s porazdelitvenimi funkcijami: vsaki porazdelitvi μ priredimo porazdelitveno funkcijo F po predpisu:

$$F(x) := \mu((-\infty, x])$$

Funkcija F je naraščajoča, z desne zvezna, njena limita v minus neskončno je enaka 0, v plus neskončno pa 1. Velja tudi obratno: vsaka taka funkcija določa porazdelitev na $\mathbb{R}$. Pripomnimo naj še, da je F nezvezna v kvečjemu števno mnogo točkah.

S porazdelitvenimi funkcijami lahko karakteriziramo tudi šibko konvergenco.

Izrek A.3.1. *Naj bodo μ in μ_n, $n \in \mathbb{N}$, porazdelitve na $\mathbb{R}$ s pripadajočimi porazdelitvenimi funkcijami F in F_n. Tedaj zaporedje μ_n šibko konvergira proti μ natanko tedaj, ko zaporedje $F_n(x)$ konvergira proti $F(x)$ za vsako točko x, v kateri je funkcija F zvezna.*

Skica dokaza. Naj μ_n šibko konvergira proti μ. Tedaj iz posledice A.2.2 takoj sledi, da zaporedje $(F_n)(x)$ konvergira k $F(x)$ v vsaki točki x, v kateri je F zvezna. Dokažimo še obratno. Naj bo $\mu_\infty := \mu$ in $F_\infty := F$. Interval $[0, 1]$, opremljen z Borelovo σ-algebro in Lebesguovo mero, je verjetnostni prostor. Za $n \in \mathbb{N} \cup \{\infty\}$ definirajmo funkcije $X_n \colon [0, 1] \to \mathbb{R}$ po predpisu:

$$X_n(\omega) := \sup\{x \mid F_n(x) < \omega\}$$

Z nekaj računanja se da dokazati, da velja $[X_n \le c] = [0, F_n(c)]$. Drugače zapisano, velja $\mathbf{P}[X_n \le c] = F_n(c)$. Za vsak n ima torej slučajna spremenljivka X_n porazdelitev μ_n. Da se tudi dokazati, da slučajne spremenljivke X_n skoraj gotovo konvergirajo proti X_∞. Potem pa mora po trditvi A.2.5 tudi zaporedje porazdelitev μ_n šibko konvergirati proti

$\mu_\infty = \mu$. Tehnične podrobnosti si lahko bralec pogleda v [21] na straneh 212–213 (lema 84.5). ■

Definicija. *Metrika Kolmogorova* na prostoru vseh porazdelitev na $\mathbb{R}$ je definirana po predpisu:

$$d_{\mathrm{K}}(\mu, \nu) := \sup_{x \in \mathbb{R}} \left| \nu((-\infty, x]) - \mu((-\infty, x]) \right|$$

To je torej metrika iz formule (A.0.1), pri čemer za testne funkcije vzamemo vse indikatorje intervalov oblike $(-\infty, x]$. Ker ti intervali tvorijo π-sistem, ki generira Borelovo σ-algebro, je po trditvi A.0.1 to res metrika. Topologija, ki jo porodi, je šibkejša od tiste, ki jo porodi metrika totalne variance, saj imamo opravka z manjšim razredom testnih funkcij.

Če zaporedje porazdelitev $(\mu_n)_{n \in \mathbb{N}}$ v metriki Kolmogorova konvergira proti porazdelitvi μ, zaporedje pripadajočih porazdelitvenih funkcij gotovo po točkah konvergira k porazdelitveni funkciji porazdelitve μ. Po izreku A.3.1 potem zaporedje μ_n šibko konvergira k μ. Iz te ugotovitve potem sledi naslednja trditev.

Trditev A.3.2. *Topologija, ki jo porodi metrika Kolmogorova, je močnejša od šibke.*

Dokaz. Naj bo A zaprta v šibki topologiji. Dokazati moramo, da je A zaprta tudi v metriki Kolmogorova. Dovolj je pokazati, da za vsako zaporedje μ_n porazdelitev iz A, ki v metriki Kolmogorova konvergira k μ, velja, da je tudi $\mu \in A$. Toda to zaporedje konvergira tudi v šibki topologiji. Ker je A v tej topologiji zaprta, mora biti res $\mu \in A$. ■

Opomba. Na prostoru $\mathrm{Pr}(\mathbb{R})$ pa obstaja tudi metrika, podobna metriki Kolmogorova, ki porodi *natanšno* šibko topologijo. To je *Lévyjeva metrika*, ki je definirana po predpisu:

$$d_{\mathrm{L}}(\mu, \nu) := \inf \left\{ \varepsilon \mid \mu((-\infty, x - \varepsilon]) - \varepsilon \le \nu((-\infty, x]) \le \mu((x + \varepsilon]) + \varepsilon \right\}$$

A.4 Wassersteinova metrika

Pri Wassersteinovi metriki vzamemo za testne funkcije vse neraztezne funkcije, t. j. Lipschitzove funkcije s konstanto 1. Metrika je definirana na prostoru vseh porazdelitev, ki so v L^1, t. j. porazdelitev, pri katerih je razdalja od izhodiščne točke v L^1.

Definicija. Naj bo (S, d) metrični prostor z Borelovo σ-algebro $\mathcal{S}$. Porazdelitev μ na $(S, \mathcal{S})$ *pripada razredu* L^1 (glede na metriko d), če za kako (vsako) točko $x \in S$ velja:

$$\int d(x, y) \mu(dy) < \infty$$

Za vsako porazdelitev μ iz L^1 in vsako enakomerno Lipschitzovo funkcijo f je potem $\int |f| \, d\mu < \infty$. Na prostoru vseh porazdelitev iz L^1 definiramo Wassersteinovo metriko d_{W} po predpisu:

$$d_{\mathrm{W}}(\mu, \nu) := \sup_{f \in \mathcal{L}} \left| \int f \, d\nu - \int f \, d\mu \right|$$

kjer je $\mathcal{L}$ razred vseh neraztenznih funkcij iz (S, d) v $\mathbb{R}$.

Opomba. Če zaporedje porazdelitev $(\mu_n)_{n\in\mathbb{N}}$ v Wassersteinovi metriki konvergira proti μ, tudi za vsako enakomerno Lipschitzovo funkcijo zaporedje $\int f\, d\mu_n$ konvergira proti $\int f\, d\mu$.

Trditev A.4.1. *Topologija, ki jo porodi Wassersteinova metrika, je močnejša od zožitve šibke topologije na* L^1.

DOKAZ. Podobno kot pri metriki Kolmogorova opazimo, da je dovolj pokazati, da vsako zaporedje $(\mu_n)_{n\in\mathbb{N}}$, ki konvergira v Wassersteinovi metriki, konvergira tudi šibko. Po izreku A.2.1 je dovolj dokazati, da za vsako zaprto množico F velja $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$. To pa dokažemo tako kot v dokazu implikacije (1) $\Rightarrow$ (2) v tem izreku, saj funkcije f_n niso le zvezne, ampak tudi enakomerno Lipschitzove. ∎

Dodatek B

Markovski procesi

V tem dodatku bomo podali osnove teorije markovskih procesov, ki je ozadje Chen–Steinove metode. Glavna vez med njimi in Chen–Steinovo metodo so seveda operatorske polgrupe. Teorija markovskih procesov je drugače lepo podana npr. v [21], teorija operatorskih polgrup pa v [10].

B.1 Prehodne porazdelitve in jedra

Prehodne porazdelitve nam povedo, kako je porazdeljena slučajna spremenljivka Y pri dani vrednosti vrednosti slučajne spremenljivke X. So ključnega pomena pri preučevanju odvisnosti slučajnih spremenljivk. Nepogrešljive so tudi v teoriji markovskih procesov, kjer preučujemo, kako se proces razvija pri dani začetni vrednosti. V resnici največkrat preučujemo prav slednje, brez konkretnega slučajnega procesa.

Definicija. Naj bo $(\Omega, \mathcal{F}, P)$ verjetnostni prostor in $A \in \mathcal{F}$. Naj bo $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ slučajna spremenljivka. *Prehodna verjetnost* dogodka B glede na X je regresija njegovega indikatorja I_B. To je torej merljiva funkcija $f\colon S \to \mathbb{R}$, ki zadošča plavzibilni enačbi:

$$f(x) = \mathbf{P}(B \mid X = x)$$

ki v resnici pomeni, da za vsako množico $A \in \mathcal{S}$ velja:

$$\int_A f \, d\mathcal{L}(X) = \mathbf{P}([X \in A] \cap B)$$

kjer je $\mathcal{L}(X)$ porazdelitev slučajne spremenljivke X.

Definicija. Naj bo $(\Omega, \mathcal{F}, \mathbf{P})$ verjetnostni prostor in naj bosta $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ slučajni spremenljivki. Funkcija $q\colon S \times \mathcal{T} \to [0, 1]$ je *prehodna porazdelitev* slučajne spremenljivke Y glede na X, če velja:

(1) Za vsak $B \in \mathcal{T}$ je funkcija $q^B\colon S \to [0, 1]$, definirana po predpisu $q^B(x) := q(x, B)$, prehodna verjetnost dogodka $[Y \in B]$ glede na X.

(2) Za vsak $x \in S$ je funkcija $q_x\colon \mathcal{T} \to [0, 1]$, definirana po predpisu $q_x(B) := q(x, B)$, verjetnostna mera na $(T, \mathcal{T})$.

Taki funkciji pravimo *prehodna porazdelitev* slučajne spremenljivke Y glede na X. To je torej usklajena družina prehodnih verjetnosti dogodkov $[Y \in B]$.

Opomba. Prehodne porazdelitve so v tesni povezavi z bolj znanimi *pogojnimi porazdelitvami glede na dano σ-algebro $\mathcal{G}$*. Če je namreč q prehodna porazdelitev slučajne spremenljivke Y glede na X, je funkcija Q, definirana po predpisu $Q(\omega, B) := q(X(\omega), B)$, pogojna porazdelitev slučajne spremenljivke Y glede na $\sigma(X)$. No, v resnici so pogojne porazdelitve poseben primer prehodnih. Če je namreč $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ slučajna spremenljivka in $\mathcal{G} \subset \mathcal{F}$, je pogojna porazdelitev slučajne spremenljivke Y glede na $\mathcal{G}$ njena prehodna porazdelitev glede na identiteto $X\colon (\Omega, \mathcal{F}) \to (\Omega, \mathcal{G})$. Več o pogojnih porazdelitvah glej npr. [9].

Včasih pa nas bodo zanimale le funkcije, ki *bi bile lahko* prehodne porazdelitve. Takim funkcijam bomo rekli *jedra*.

Definicija. Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora. Funkcija $q\colon S \times \mathcal{T} \to [0, 1]$ je *jedro* iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$, če izpolnjuje naslednja dva pogoja, ki ustrezata pogojema pri definiciji prehodne porazdelitve:

(1) Za vsak $B \in \mathcal{T}$ je funkcija q^B, definirana po predpisu $q^B(x) := q(x, B)$, $\mathcal{S}$-merljiva.

(2) Za vsak $x \in S$ je funkcija q_x, definirana po predpisu $q_x(B) := q(x, B)$, verjetnostna mera na $(T, \mathcal{T})$.

Jedro na prostoru $(S, \mathcal{S})$ je jedro iz $(S, \mathcal{S})$ v $(S, \mathcal{S})$.

Če hočemo dokazati, da je q prehodna porazdelitev slučajne spremenljivke Y glede na X, moramo torej preveriti, da je q jedro iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ in da velja:

$$\int_A q^B \, d\mathcal{L}(X) = \mathbf{P}[X \in A, Y \in B] \tag{B.1.1}$$

Opomba. Tako kot je regresija poseben primer pogojnega matematičnega upanja, so tudi pogojne porazdelitve glede na dano σ-algebro $\mathcal{G} \subset \mathcal{F}$ poseben primer prehodnih porazdelitev. Za X namreč vzamemo identiteto $(\Omega, \mathcal{F}) \to (\Omega, \mathcal{G})$.

Seveda bi bilo o prehodnih porazdelitvah nesmiselno govoriti, če ne bi obstajale. Na srečo pa obstajajo za dovolj velik razred merljivih prostorov. Prehodna porazdelitev obstaja, brž ko je prostor $(T, \mathcal{T})$ Borelov, to je takrat, ko je kot merljiv prostor izomorfen kaki Borelovi podmnožici realne osi (z Borelovo σ-algebro). Konstrukcija prehodnih porazdelitev je povsem enaka konstrukciji pogojnih porazdelitev glede na dano σ-algebro, ki so njihov poseben primer (konstrukcijo pogojnih porazdelitev si bralec lahko pogleda v [9] na strani 230, izrek 1.6). Razred Borelovih prostorov je kar velik, saj je znano, da je vsak poln separabilen metrični prostor Borelov (glej [9], stran 33, izrek 4.12).

B.2 Prehodne porazdelitve in preslikave

V tem razdelku bomo preučili razne lastnosti slučajne spremenljivke $f(X, Y)$, kot so matematično upanje, regresija glede na X in prehodna porazdelitev glede na X, pri dani prehodni porazdelitvi slučajne spremenljivke Y glede na X. Ključni korak je izražava regresije.

Izrek B.2.1. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora in naj bo $f\colon S \times T \to \mathbb{R}$ merljiva funkcija. Dano naj bo jedro q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$. Definirajmo funkcijo $Pf\colon D_f \to \mathbb{R}$ po predpisu:*

$$(Pf)(x) := \int f(x, y)\, q(x, dy) \tag{B.2.1}$$

Tu je D_f množica vseh tistih $x \in S$, za katere je $\int |f(x, y)|\, q(x, dy) < \infty$. Množica D_f je merljiva in tudi funkcija Pf je merljiva.

Naj bosta $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ slučajni spremenljivki in naj bo q prehodna porazdelitev slučajne spremenljivke Y glede na X. Naslednji trditvi sta ekvivalentni:

(1) *$f \in L^1(\mathcal{L}(X, Y))$ (oz. $f(X, Y) \in L^1(\mathbf{P})$).*

(2) *Velja $\mathcal{L}(X)(D_f) = 1$ in $P|f| \in L^1(\mathcal{L}(X))$ (če ima $D_f = D_{|f|}$ mero 1, funkcija $P|f|$ že določa ekvivalenčni razred funkcij glede na enakost skoraj povsod glede na $\mathcal{L}(X)$).*

V primeru, ko trditvi veljata, je Pf regresija slučajne spremenljivke $f(X, Y)$ glede na X. Funkcija Pf je tudi edina funkcija, ki je dobra za vse možne ustrezne pare (X, Y).

Skica dokaza. Izrek najprej dokažemo za primer, ko je f indikator pravokotnika. S pomočjo Dynkinove leme, znane tudi kot izrek π–λ, izrek posplošimo na primer, ko je f indikator poljubne množice iz produktne σ-algebre $\mathcal{S} \times \mathcal{T}$. Končno z uporabo linearnosti, izreka o monotoni konvergenci in razcepa funkcije na pozitivni in negativni del izrek dokažemo tudi za poljubno merljivo funkcijo. Dokaz izreka v vseh podrobnostih ima sicer kar nekaj tehničnih sitnosti, vendar nikjer ni potreben prav globok premislek. ■

Posledica B.2.2. *Naj bodo f, X in Y in q kot v izreku B.2.1. Če bodisi držita trditvi (1) in (2), bodisi je $f \geq 0$, velja:*

$$\mathbf{E}\big(f(X, Y)\big) = \int \int f(x, y)\, q(x, dy)\, \mathcal{L}(X)(dx) \tag{B.2.2}$$

Dokaz. V primeru, ko držita trditvi (1) in (2), vse skupaj takoj sledi iz izreka. Če pa je f nenegativna, je f limita naraščajočega zaporedja omejenih merljivih funkcij, za katere trditvi (1) in (2) gotovo držita. Potem pa enačbi sledita iz izreka o monotoni konvergenci. ■

Opomba. Formula B.2.2 ne pomeni nič drugega kot:

$$\mathbf{E}\big(f(X, Y)\big) = \mathbf{E}\Big(\mathbf{E}\big(f(X, Y) \mid X\big)\Big)$$

Opomba. V primeru, ko sta slučajni spremenljivki X in Y neodvisni, lahko postavimo kar $q(x, B) = \mathbf{P}[Y \in B]$ za vsak x in potem formula (B.2.2) ni čisto nič drugega kot Fubinijev izrek.

Izrek B.2.1 nam izraža regresijo slučajne spremenljivke $f(X,Y)$ glede na X. Z njegovo pomočjo pa lahko rekonstruiramo kar celo prehodno porazdelitev te slučajne spremenljivke. Naj bosta torej $X\colon (\Omega,\mathcal{F}) \to (S,\mathcal{S})$ in $Y\colon (\Omega,\mathcal{F}) \to (T,\mathcal{T})$ slučajni spremenljivki in $f\colon (S\times T, \mathcal{S}\times\mathcal{T}) \to (U,\mathcal{U})$ merljiva preslikava (zdaj ni več nujno, da je $U = \mathbb{R}$). Naj bo $C \in \mathcal{U}$. Po izreku B.2.1 je prehodna verjetnost dogodka $[f(X,Y)\in C]$, torej regresija slučajne spremenljivke $I_C(f(X,Y))$ glede na X, definirana po predpisu:

$$g(x) := \int I_C(f(x,y))\, q(x,dy) = \int I_{f_x^{-1}(C)}(y)\, q(x,dy) = q\big(x, f_x^{-1}(C)\big)$$

in to je tudi edina regresija, ki je dobra za vse možne pare (X,Y), pri katerih je dano jedro q prehodna porazdelitev slučajne spremenljivke Y glede na X. Označili smo $f_x(y) = f(x,y)$.

Izrek B.2.3. *Naj bodo $(S,\mathcal{S})$, $(T,\mathcal{T})$ in $(U,\mathcal{U})$ merljivi prostori in naj bo q jedro iz $(S,\mathcal{S})$ v $(T,\mathcal{T})$. Dana naj bo še merljiva preslikava $f\colon (S\times T,\mathcal{S}\times\mathcal{T}) \to (U,\mathcal{U})$. Tedaj obstaja natanko eno jedro q' iz $(S,\mathcal{S})$ v $(U,\mathcal{U})$, za katerega velja: če sta $X\colon(\Omega,\mathcal{F}) \to (S,\mathcal{S})$ in $Y\colon (\Omega,\mathcal{F}) \to (T,\mathcal{T})$ poljubni slučajni spremenljivki in je q prehodna porazdelitev slučajne spremenljivke Y glede na X, je q' prehodna porazdelitev slučajne spremenljivke $f(X,Y)$ glede na X. Velja formula:*

$$q'(x,C) := q\big(x, f_x^{-1}(C)\big) \tag{B.2.3}$$

Dokaz. Enoličnost smo že dokazali. Iz izreka B.2.1 sledi, da je $(q')^C$ merljiva za vsako množico $C \in \mathcal{S}\times\mathcal{T}$, očitno pa je tudi $C \mapsto q'(x,C)$ verjetnostna mera za vsak $x\in S$. Torej je q' jedro. Če je q prehodna porazdelitev slučajne spremenljivke Y glede na X, smo tudi že dokazali, da je za vsak $C\in\mathcal{U}$ funkcija $x\mapsto q(x,C)$ prehodna verjetnost dogodka $[f(X,Y)\in C]$ glede na X. To pa je vse, kar je bilo treba dokazati. ■

B.3 Jedra kot transformacije funkcij

Porazdelitev slučajne spremenljivke Y je po definiciji preslikava $B \mapsto \mathbf{P}[Y\in B]$, ki je verjetnostna mera. Lahko pa jo opišemo tudi s preslikavo $f \mapsto \mathbf{E}(f(Y))$. Vsaki verjetnostni meri μ na merljivem prostoru $(T,\mathcal{T})$ torej pripada funkcional $\Phi\colon \flat(T,\mathcal{T}) \to \mathbb{R}$, ki deluje po predpisu:

$$\Phi(f) := \int f\, d\mu \tag{B.3.1}$$

S $\flat(T,\mathcal{T})$ bomo označevali prostor vseh omejenih merljivih funkcij iz T v $\mathbb{R}$. Funkcional Φ ima nekaj očitnih lastnosti:

- Linearnost: velja $\Phi(\alpha f + \beta g) = \alpha\Phi(f) + \beta\Phi(g)$.
- Pozitivnost: če je $f \geq 0$, je tudi $\Phi(f) \geq 0$.
- Monotona konvergenca: če zaporedje nenegativnih funkcij f_n po točkah narašča proti omejeni funkciji f, je tudi $\lim_{n\to\infty}\Phi(f_n) = \Phi(f)$.

- Normaliziranost: velja $\Phi(1) = 1$.

Definicija. *Kontrakcija* je linearna preslikava med normiranima prostoroma, ki ima normo največ 1.

Trditev B.3.1. *Vsak linearni funkcional Φ na $\flat(T, \mathcal{T})$, ki je poleg tega še pozitiven in v zgornjem smislu normaliziran, je tudi kontrakcija.*

Dokaz. Funkciji $\|f\| + f$ in $\|f\| - f$ sta očitno pozitivni. Potem pa velja:

$$\Phi(f) = \|f\| - \Phi(\|f\| - f) \leq \|f\|$$
$$\Phi(f) = \Phi(\|f\| + f) - \|f\| \geq -\|f\|$$

Od tod pa že sledi, da je Φ kontrakcija. ∎

Funkcional Φ nam popolnoma opiše mero μ, saj iz enačbe (B.3.1) sledi:

$$\mu(B) := \Phi(I_B) \tag{B.3.2}$$

kjer I_B pomeni indikator množice $B \in \mathcal{T}$ (to je funkcija, ki je na B enaka 1, drugje pa 0). Velja pa še več: preslikava, ki verjetnostni meri priredi ustrezen funkcional, se da obrniti.

Trditev B.3.2. *Naj bo $(T, \mathcal{T})$ merljiv prostor in Φ realni funkcional na $\flat(T, \mathcal{T})$, ki ima zgoraj opisane lastnosti. Tedaj obstaja natanko ena verjetnostna mera μ na $(T, \mathcal{T})$, za katero velja enačba (B.3.1). Mera μ je podana z enačbo (B.3.2).*

Dokaz. Enoličnost smo že dokazali. Če za podani funkcional Φ definiramo preslikavo μ po enačbi (B.3.2), iz zgornjih lastnosti sledi, da je μ res verjetnostna mera na $(T, \mathcal{T})$. Preveriti je treba le še enačbo (B.3.1). Vemo, da enačba velja za indikatorje. Iz linearnosti sledi, da velja tudi za merljive enostavne funkcije, iz monotone konvergence, da velja za vse pozitivne omejene merljive funkcije, in spet iz linearnosti, da velja za vse omejene merljive funkcije. ∎

Govoriti o verjetnostnih merah na merljivem prostoru $(T, \mathcal{T})$ je torej isto kot govoriti o funkcionalih na $\flat(T, \mathcal{T})$ s prej naštetimi lastnostmi.

Porazdelitev dane slučajne spremenljivke Y torej lahko opišemo s preslikavo $f \mapsto \mathbf{E}(f(Y))$. Nastane vprašanje, ali lahko vse to naredimo tudi pogojno na dano slučajno spremenljivko X: prehodno porazdelitev slučajne spremenljivke Y glede na X bi radi opisali s preslikavo, ki omejeni merljivi funkciji f priredi regresijo slučajne spremenljivke $f(Y)$ glede na X. Zdaj torej ne bomo več dobili linearnega funkcionala, ampak linearni operator, ki bo funkciji na $(T, \mathcal{T})$ priredil funkcijo na $(S, \mathcal{S})$.

Porazdelitev je po definiciji verjetnostna mera, prehodna porazdelitev pa je jedro. Naša naloga je torej jedra opisati s primernimi operatorji. Ključ do tega je naslednja posledica izreka B.2.1.

Posledica B.3.3. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora in naj bo $f \colon T \to \mathbb{R}$ omejena merljiva funkcija. Dano naj bo jedro q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$. Definirajmo funkcijo $Pf \colon S \to \mathbb{R}$ po predpisu:*

$$(Pf)(x) := \int f(y)\, q(x, dy) \tag{B.3.3}$$

Funkcija Pf je merljiva. Če sta X: $(\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in Y: $(\Omega, \mathcal{F}) \to (T, \mathcal{T})$ slučajni spremenljivki in je q prehodna porazdelitev slučajne spremenljivke Y glede na X, je Pf regresija slučajne spremenljivke $f(Y)$ glede na X. Funkcija Pf je tudi edina funkcija, ki je dobra za vse možne ustrezne pare (X, Y).

Jedru q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ smo torej priredili operator P: $\flat(T, \mathcal{T}) \to \flat(S, \mathcal{S})$. Le-ta ima spet naslednje očitne lastnosti:

- Linearnost: velja $P(\alpha f + \beta g) = \alpha Pf + \beta Pg$.
- Pozitivnost: če je $f \geq 0$, je tudi $Pf \geq 0$.
- Monotona konvergenca: če zaporedje nenegativnih funkcij f_n po točkah narašča proti omejeni funkciji f, tudi zaporedje Pf_n po točkah narašča k Pf.
- Normaliziranost: velja $P(1) = 1$.

Opomba B.3.4. *Vsak linearni operator P iz $\flat(T, \mathcal{T})$ v prostor merljivih funkcij na $(S, \mathcal{S})$, ki je poleg tega še pozitiven in v zgornjem smislu normaliziran, je prav tako tudi kontrakcija: velja $\|P(f)\| \leq \|f\|$.* Za vsak $x \in \mathcal{S}$ je namreč funkcional $f \mapsto (Pf)(x)$ linearen, pozitiven in normaliziran, torej je po opombi B.3.1 kontrakcija, od koder že sledi, da je tudi P kontrakcija.

Vsakemu jedru smo torej priredili linearni operator iz $\flat(T, \mathcal{T})$ v $\flat(S, \mathcal{S})$. Podobno kot prej je tudi tu mogoče prireditev, ki smo jo pravkar opisali, obrniti.

Izrek B.3.5. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora in P operator iz $\flat(T, \mathcal{T})$ v $(S, \mathcal{S})$, ki ima zgoraj opisane lastnosti. Tedaj obstaja natanko eno jedro q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$, za katerega velja enačba (B.3.3). Jedro q je podano z enačbo:*

$$q(x, B) = (PI_B)(x) \tag{B.3.4}$$

Jedro q je tudi edino jedro, za katerega velja: če sta X: $(\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in Y: $(\Omega, \mathcal{F}) \to (T, \mathcal{T})$ slučajni spremenljivki in je Pf regresija slučajne spremenljivke $f(Y)$ glede na X za vsako funkcijo $f \in \flat(T, \mathcal{T})$, je q prehodna porazdelitev slučajne spremenljivke Y glede na X.

Dokaz. Da mora biti jedro q res podano z enačbo (B.3.4), sledi iz enačbe (B.3.3). Enoličnost smo torej že dokazali. Če za podani funkcional Φ definiramo preslikavo q po enačbi (B.3.4), iz zgornjih lastnosti sledi, da je q res jedro iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$. Preveriti je treba le še enačbo (B.3.3). Vemo, da enačba velja za indikatorje. Podobno kot v dokazu trditve B.3.2 tudi tu prav lahko dokažemo, da v resnici velja za vse funkcije iz $\flat(T, \mathcal{T})$.

Dokažimo še drugi del izreka. Enoličnost dokažemo tako, da za vsak $x \in S$ izberemo taki slučajni spremenljivki X in Y, da je X skoncentrirana v x in da velja $E(f(Y)) = (Pf)(x)$ za vsako funkcijo $f \in \flat(T, \mathcal{T})$. Po trditvi B.3.2 namreč obstaja taka verjetnostna mera μ na $(T, \mathcal{T})$, da za vsako funkcijo $f \in \flat(T, \mathcal{T})$ velja $(Pf)(x) = \int f\, d\mu$, potem pa X in Y definiramo na verjetnostnem prostoru $(T, \mathcal{T}, \mu)$ in za X vzamemo konstanto x, za Y pa identiteto na T. Potem je Pf gotovo regresija slučajne spremenljivke $f(Y)$ glede na X, torej je tudi PI_C prehodna verjetnost dogodka $[Y \in C]$ glede na X. Le-ta je v x natančno določena, torej mora biti res $q(x, C) = (PI_C)(x)$. Če q definiramo po tem predpisu in sta X in Y poljubni taki slučajni spremenljivki, da je Pf regresija slučajne spremenljivke $f(Y)$ glede na X, je potem očitno q prehodna porazdelitev slučajne spremenljivke Y glede na X in izrek je dokazan. ■

Govoriti o jedrih iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ je torej isto kot govoriti o operatorjih iz $\flat(T, \mathcal{T})$ v $(S, \mathcal{S})$ s prej naštetimi lastnostmi.

B.4 Jedra kot transformacije porazdelitev

Naj bosta X in Y slučajni spremenljivki in naj bo ν porazdelitev slučajne spremenljivke X, q pa prehodna porazdelitev slučajne spremenljivke Y glede na X. Zanima nas porazdelitev slučajne spremenljivke Y, ki jo bomo označili s $\mathbf{P}_\nu$. Formula (B.1.1) nam pove:

$$\mathbf{P}_\nu(B) = \int q(x, B)\, \nu(dx) \tag{B.4.1}$$

Ni težko preveriti, da je za vsako porazdelitev ν na danem merljivem prostoru $(S, \mathcal{S})$ in vsako jedro q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ funkcija $\mathbf{P}_\nu$ porazdelitev na $(T, \mathcal{T})$. Na ta način smo vsakemu jedru iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ priredili transformacijo, ki vsaki porazdelitvi na $(S, \mathcal{S})$ priredi porazdelitev na $(T, \mathcal{T})$.

Porazdelitvi $\mathbf{P}_\nu$ seveda pripada matematično upanje $\mathbf{E}_\nu$. Za $x \in S$ označimo še:

$$\begin{aligned} \mathbf{P}_x(B) &:= \mathbf{P}_{\delta_x}(B) = q(x, B) \\ \mathbf{E}_x(f) &:= \mathbf{E}_{\delta_x}(B) = \int f(y)\, q(x, dy) = (Pf)(x) \end{aligned}$$

Tu je δ_x Diracova mera, P pa je operator, ki pripada jedru q po oznakah iz prejšnjega razdelka. Iz zgornjih zvez in izsledkov prejšnjega razdelka je tudi jasno, da je za vsak $B \in \mathcal{T}$ funkcija $x \mapsto \mathbf{P}_x(B)$ merljiva, prav tako pa je tudi za vsak $f \in \flat(T, \mathcal{T})$ merljiva funkcija $x \mapsto \mathbf{E}_x(f)$.

Za dano množico $B \in \mathcal{T}$ ali funkcijo $f \in \flat(T, \mathcal{T})$ ter slučajno spremenljivko $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ lahko potem definiramo realni slučajni spremenljivki $\mathbf{P}_X(B)$ in $\mathbf{E}_X(f)$ po očitnem predpisu:

$$\begin{aligned} \mathbf{P}_X(B)(\omega) &:= \mathbf{P}_{X(\omega)}(B) \\ \mathbf{E}_X(f)(\omega) &:= \mathbf{E}_{X(\omega)}(f) \end{aligned}$$

Kdaj je dano jedro q prehodna porazdelitev slučajne spremenljivke Y glede na X, lahko zdaj zapišemo z našimi novimi oznakami. Iz posledice B.3.3 in izreka B.3.5 takoj sledi naslednja trditev.

Trditev B.4.1. *Naj bo q jedro iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$. Dani naj bosta še slučajni spremenljivki $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$. Naslednje trditve so ekvivalentne:*

(1) *Jedro q je prehodna porazdelitev slučajne spremenljivke Y glede na X.*

(2) *Za vsak $B \in \mathcal{T}$ velja:*

$$\mathbf{P}_X(B) = \mathbf{P}\big([Y \in B] \mid X\big)$$

(3) *Za vsak* $f \in \flat(T, \mathcal{T})$ *velja:*

$$\mathbf{E}_X(f) = \mathbf{E}\big(f(Y) \,\big|\, X\big)$$

Nastane vprašanje, ali se da preslikava, ki jedru priredi ustrezno transformacijo, obrniti. Za primerno transformacijo $\nu \mapsto \mathbf{P}_\nu$ bi radi poiskali tako jedro q, da bo veljala enačba (B.4.1). Najprej si oglejmo primer, ko je ν Diracova mera δ_x, $x \in S$. V tem primeru iz enačbe (B.4.1) dobimo, da mora biti jedro q definirano po predpisu:

$$q(x, B) := \mathbf{P}_x(B) \tag{B.4.2}$$

Potreben in zadosten pogoj, da bo veljala enačba (B.4.1), je očitno enak, da je za vsak $B \in \mathcal{T}$ funkcija $x \mapsto \mathbf{P}_x(B)$ merljiva in velja:

$$\mathbf{P}_\nu(B) = \int \mathbf{P}_x(B)\, \nu(dx) \tag{B.4.3}$$

Za take transformacije bomo rekli, da so *usklajene z integracijo*.

Vsakemu jedru q torej po enačbi (B.4.1) pripada transformacija $F\colon \Pr(S, \mathcal{S}) \to \to \Pr(T, \mathcal{T})$, $\nu \mapsto \mathbf{P}_\nu$, ki je usklajena z integracijo. Velja tudi obratno: vsaki transformaciji porazdelitev $\nu \mapsto \mathbf{P}_\nu$, ki je usklajena z integracijo, po formuli (B.4.2) pripada jedro q in za to jedro velja tudi enačba (B.4.1). S $\Pr(S, \mathcal{S})$ bomo označevali prostor vseh verjetnostnih mer na danem merljivem prostoru $(S, \mathcal{S})$.

V zvezi s transformacijami porazdelitev, ki so usklajene z integracijo, omenimo še trditev, ki jo bomo potrebovali kasneje.

Trditev B.4.2. *Kompozitum transformacij, ki sta usklajeni z integracijo, je tudi usklajen z integracijo.*

Dokaz. Naj bosta transformaciji $F\colon \Pr(S, \mathcal{S}) \to \Pr(T, \mathcal{T})$ in $G\colon \Pr(T, \mathcal{T}) \to (U, \mathcal{U})$ usklajeni z integracijo. Za vsak $B \in \mathcal{T}$ je torej funkcija $x \mapsto F(\delta_x)(B)$ merljiva in za vsako porazdelitev $\nu \in \Pr(S, \mathcal{S})$ velja:

$$F(\nu)(B) = \int F(\delta_x)(B)\, \nu(dx)$$

Iz linearnosti in izreka o dominirani konvergenci sledi, da je potem tudi za vsako funkcijo $g \in \flat(T, \mathcal{T})$ funkcija $x \mapsto \int g(y)\, F(\delta_x)(dy)$ merljiva in velja:

$$\int g(y)\, F(\nu)(dy) = \int \int g(y)\, F(\delta_x)(dy)\, \nu(dx)$$

Potem pa velja:

$$\begin{aligned} G(F(\nu))(C) &= \int G(\delta_y)(C)\, F(\nu)(dy) = \int \int G(\delta_y)\, F(\delta_x)(y)\, \nu(dx) = \\ &= \int G(F(\delta_x))\, \nu(dx) \end{aligned}$$

Od tod pa sledi vse, kar je bilo treba dokazati. ■

Spoznali smo, da je govoriti o jedrih isto kot govoriti o določenih transformacijah porazdelitev, podobno kot smo v prejšnjem razdelku spoznali, da je to isto kot govoriti o primernih linearnih operatorjih. Vendar pa je vsa zgodba račun brez krčmarja, če se ugotovljene ekvivalence med našimi objekti ne odražajo na slučajnih spremenljivkah.

Predpis, ki jedru priredi transformacijo $\nu \mapsto \mathbf{P}_\nu$, se na slučajnih spremenljivkah lepo odraža: če ima slučajna spremenljivka X porazdelitev ν in je q prehodna porazdelitev slučajne spremenljivke Y glede na X, je $\mathbf{P}_\nu$ porazdelitev slučajne spremenljivke Y. Za obratni predpis pa bi želeli tole: če ima X porazdelitev ν, Y pa porazdelitev $\mathbf{P}_\nu$, naj bi bilo jedro q, dobljeno po formuli (B.4.2), prehodna porazdelitev slučajne spremenljivke Y glede na X. Vendar pa to še zdaleč ni res. Naj bo na primer $S = T = \{0, 1\}$, za σ-algebri vzamemo potenčni množici, ν naj bo enakomerna porazdelitev na S, jedro q pa naj bo deterministično: za $x \in S$ naj bo $q(x, B) := I_B(x)$. Tedaj je tudi $\mathbf{P}_\nu$ enakomerna porazdelitev na T.

Jedro q zahteva, da morata biti slučajni spremenljivki X in Y skoraj gotovo enaki. To pa seveda ni res, če za X in Y vzamemo neodvisni slučajni spremenljivki, ki sta enakomerno porazdeljeni na $\{0, 1\}$. Obratni predpis se torej ne odraža na slučajnih spremenljivkah, tako kot bi želeli.

Da se bo tudi obratni predpis lepo odražal, bomo morali zahtevati dodatno strukturo slučajnih spremenljivk in jeder. Sprejeli bomo precej hudo zahtevo, da mora biti X deterministično odvisna od Y. Obstajati bo torej morala taka merljiva (navadno surjektivna) preslikava $p\colon (T, \mathcal{T}) \to (S, \mathcal{S})$, da bo veljalo $X = p(Y)$. Preslikavi p bomo rekli *projekcija*.

Omejitev se zdi huda, vendar bo pri vseh stvareh, ki jih bomo preučavali, pogoj, da je $X = p(Y)$, v resnici izpolnjen. Slučajna spremenljivka X bo namreč opisovala začetno stanje, Y pa cel razvoj slučajnega procesa. Zanimalo nas bo, kako se proces razvija pri danem začetnem stanju. Ker podatek o celem razvoju vsebuje tudi podatek o začetnem stanju, bomo projekcijo p vedno lahko našli.

Tudi primer, ko gledamo prehodno porazdelitev poljubne slučajne spremenljivke Y glede na X, lahko prevedemo na primer, ko obstaja naravna projekcija p: gledamo namreč porazdelitev slučajnega vektorja (X, Y) glede na X, potem pa lahko za p vzamemo projekcijo na prvo koordinato. Omejitev, ki jo bomo zahtevali, torej sploh ni tako huda, kot se zdi.

Omejitev, ki jo zahtevamo za slučajne spremenljivke, moramo seveda zahtevati tudi za jedra. Kakšno naj bo dobro jedro q, ki je prehodna porazdelitev slučajne spremenljivke Y glede na X, pri čemer velja $X = p(Y)$? Če je $X = x$, je seveda $p(Y) = x$, torej je $Y \in p^{-1}(\{x\})$. V skladu s tem postavimo naslednjo definicijo.

Definicija. Jedro q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ je *usklajeno s projekcijo* $p\colon (T, \mathcal{T}) \mapsto (S, \mathcal{S})$, če je za vsak $x \in S$ verjetnostna mera $B \mapsto q(x, B)$ skoncentrirana na vlaknu $p^{-1}(\{x\})$, torej, če velja:

$$q\big(x, p^{-1}(\{x\})\big) = 1 \tag{B.4.4}$$

Zgled 1. Naj bosta $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ slučajni spremenljivki in naj bo q prehodna porazdelitev slučajne spremenljivke Y glede na X. Iz izreka B.2.3

sledi, da je funkcija q', definirano po predpisu:

$$q'(x, C) := q(x, C_x) \tag{B.4.5}$$

kjer je $C_x = \{y \in T \mid (x, y) \in T\}$, prehodna porazdelitev slučajne spremenljivke (X, Y) glede na X. Če s $p\colon S \times T \to S$ označimo projekcijo na prvo koordinato, je $X = p(X, Y)$ in tudi jedro q' je usklajeno s projekcijo p.

Nastane vprašanje, ali vedno obstaja prehodna porazdelitev, ki je usklajena z dano projekcijo. Za ilustracijo bomo navedli naslednjo trditev, ki pa je kasneje ne bomo potrebovali.

Trditev B.4.3. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ Borelova prostora. Dani naj bosta slučajni spremenljivki $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ ter merljiva surjekcija $p\colon (T, \mathcal{T}) \to (S, \mathcal{S})$ z merljivim desnim inverzom s. Naj bo $X = p(Y)$. Tedaj obstaja prehodna porazdelitev slučajne spremenljivke Y glede na X, ki je usklajena s projekcijo p.*

Dokaz. Ker je prostor $(T, \mathcal{T})$ Borelov, obstaja prehodna porazdelitev q slučajne spremenljivke Y glede na X. Seveda ni rečeno, da je jedro q usklajeno s projekcijo p. Pokazali pa bomo, da je skoraj usklajeno.

Naj bo Γ graf projekcije p z zamenjanimi koordinatami, torej:

$$\Gamma := \{(p(y), y) \mid y \in T\} = \{(x, y) \in S \times T \mid x = p(y)\}$$

Preverimo najprej, da je množica Γ merljiva. Ker je prostor $(S, \mathcal{S})$ Borelov, lahko na njem prek izomorfizma na Borelovo podmnožico realne osi definiramo odštevanje: obstaja taka merljiva preslikava $\delta\colon S \times S \to \mathbb{R}$, da je $x = y$ natanko tedaj, ko je $\delta(x, y) = 0$. Potem pa lahko zapišemo:

$$\Gamma = \{(x, y) \in S \times T \mid \delta(x, p(y)) = 0\}$$

kar je res merljiva množica.

Naj bo $f := I_\Gamma$. Ker je $X = p(Y)$, je $f(X, Y) = 1$. Po posledici B.2.2 velja:

$$\begin{aligned}
1 = \mathbf{E}\bigl(f(X, Y)\bigr) &= \int\!\!\int f(x, y)\, q(x, dy)\, \mathcal{L}(X)(dx) = \\
&= \int\!\!\int I_{p^{-1}(\{x\})}(y)\, q(x, dy)\, \mathcal{L}(X)(dx) = \\
&= \int q\bigl(x, p^{-1}(\{x\})\bigr)\, \mathcal{L}(X)(dx)
\end{aligned}$$

Sledi:

$$\int \bigl(1 - q\bigl(x, p^{-1}(\{x\})\bigr)\, \mathcal{L}(X)(dx) = 0$$

Od tod pa sledi, da je $q\bigl(x, p^{-1}(\{x\})\bigr) = 1$ za skoraj vsak $x \in S$ glede na $\mathcal{L}(X)$. Množica tistih x, kjer se to ne zgodi, je torej merljiva in ima mero 0. Označimo jo z N. Izberimo poljubno verjetnostno mero ν na $(T, \mathcal{T})$ in definirajmo:

$$q'(x, B) = \begin{cases} q(x, B) & x \notin N \\ I_B(s(x)) & x \in N \end{cases}$$

Tedaj je jedro q' še vedno prehodna porazdelitev slučajne spremenljivke Y glede na X, saj smo q popravili le na množici z mero 0, je pa tudi usklajeno s projekcijo p. ∎

Oglejmo si zdaj, kako se ugotovitve z začetka razdelka izboljšajo, če privzamemo usklajenost s projekcijo p.

Izrek B.4.4. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora in naj bo $p\colon (T, \mathcal{T}) \to (S, \mathcal{S})$ merljiva preslikava. Naj bo ν porazdelitev na $(S, \mathcal{S})$, q pa jedro iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$, usklajeno s p. Naj bo $\mathbf{P}_\nu$ porazdelitev na $(T, \mathcal{T})$, definirana po enačbi (B.4.1). Tedaj sta za poljubno slučajno spremenljivko $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ naslednji trditvi ekvivalentni:*

(1) *Slučajna spremenljivka $X := p(Y)$ ima porazdelitev ν in q je prehodna porazdelitev slučajne spremenljivke Y glede na X.*

(2) *Slučajna spremenljivka Y ima porazdelitev $\mathbf{P}_\nu$.*

Vsaka porazdelitev $\mathbf{P}'_\nu$, za katero bodisi za vsak Y velja implikacija (2) $\Rightarrow$ (1) bodisi za vsak Y velja implikacija (2) $\Rightarrow$ (1), je enaka $\mathbf{P}_\nu$. Tako smo jedru q, usklajenemu s p, priredili transformacijo $\nu \mapsto \mathbf{P}_\nu$. Ta transformacija je usklajena z integracijo.

Dokaz. Že na začetku razdelka smo dokazali, da za porazdelitev $\mathbf{P}_\nu$ za vsak Y velja implikacija (1) $\Rightarrow$ (2). Pokažimo še obratno implikacijo. Iz formule (B.4.4) sledi:

$$q(x, B) = q\big(x, p^{-1}(\{x\}) \cap B\big)$$

Sledi:

$$\begin{aligned} q\big(x, p^{-1}(A) \cap B\big) &= q\big(x, p^{-1}(\{x\}) \cap p^{-1}(A) \cap B\big) = \\ &= I_A(x)\, q\big(x, p^{-1}(\{x\}) \cap B\big) = \\ &= I_A(x)\, q(x, B) \end{aligned}$$

Naj ima zdaj Y porazdelitev $\mathbf{P}_\nu$. Pokazati moramo, da ima X porazdelitev ν in da je q prehodna porazdelitev slučajne spremenljivke Y glede na X. Prva trditev sledi iz naslednje verige enakosti:

$$\begin{aligned} \mathbf{P}[X \in A] = \mathbf{P}[p(Y) \in A] = \mathbf{P}_\nu\big(p^{-1}(A)\big) &= \int q\big(x, p^{-1}(A)\big)\, \nu(dx) = \\ &= \int_A q(x, T)\, \nu(dx) = \int_A \nu(dx) = \nu(A) \end{aligned}$$

Ključni peti enačaj sledi iz prej izpeljane verige enakosti. Slučajna spremenljivka X ima torej res porazdelitev ν. Ker je q jedro, je za drugo trditev dovolj dokazati enačbo (B.1.1). Le-ta pa sledi iz naslednje verige enakosti:

$$\begin{aligned} \int_A q(x, B) \mathcal{L}(X)(dx) = \int I_A(x)\, q(x, B)\, \nu(dx) &= \int q\big(x, p^{-1}(A) \cap B\big)\, \nu(dx) = \\ &= \mathbf{P}[Y \in p^{-1}(A) \cap B] = \mathbf{P}[X \in A, Y \in B] \end{aligned}$$

Dokazali smo, da porazdelitev $\mathbf{P}_\nu$ res izpolnjuje pogoje izreka. Da je transformacija $\nu \mapsto \mathbf{P}_\nu$ usklajena z integracijo, takoj sledi iz formule (B.4.1). Pokazati je treba le še enoličnost.

Naj bo $\mathbf{P}'_\nu$ poljubna porazdelitev na $(T, \mathcal{T})$, za katero za vsak Y velja (1) $\Rightarrow$ (2). Izberimo slučajno spremenljivko Y s porazdelitvijo $\mathbf{P}_\nu$. Tedaj velja (1), torej ima Y porazdelitev $\mathbf{P}'_\nu$. Potem pa je $\mathbf{P}'_\nu = \mathbf{P}_\nu$.

Naj bo zdaj $\mathbf{P}'_\nu$ poljubna porazdelitev, za katero za vsak Y velja implikacija (2) $\Rightarrow$ (1). Zdaj izberimo slučajno spremenljivko Y s porazdelitvijo $\mathbf{P}'_\nu$. Spet velja (1), torej ima Y porazdelitev $\mathbf{P}_\nu$, potem pa je spet $\mathbf{P}'_\nu = \mathbf{P}_\nu$. ∎

Iz tega izreka in zgleda na začetku razdelka dobimo naslednjo posledico.

Posledica B.4.5. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora. Naj bo ν porazdelitev na $(S, \mathcal{S})$, q pa jedro iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$. Definirajmo porazdelitev $\mathbf{P}_\nu$ na $(S \times T, \mathcal{S} \times \mathcal{T})$ po predpisu:*

$$\mathbf{P}_\nu(C) = \int q(x, C_x)\, \nu(dy) \tag{B.4.6}$$

(označili smo $C_x := \{y \in T \mid (x, y) \in C\}$). Tedaj sta za poljubni slučajni spremenljivki $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ in $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ naslednji trditvi ekvivalentni:

(1) *Slučajna spremenljivka X ima porazdelitev ν, q pa je prehodna porazdelitev slučajne spremenljivke Y glede na X.*

(2) *Slučajni vektor (X, Y) ima porazdelitev $\mathbf{P}_\nu$.*

Vsaka porazdelitev $\mathbf{P}'_\nu$, za katero bodisi za poljubna X in Y velja implikacija (2) $\Rightarrow$ (1) bodisi za poljubna X in Y velja implikacija (2) $\Rightarrow$ (1), je enaka $\mathbf{P}_\nu$. Tako smo jedru q priredili transformacijo $\nu \mapsto \mathbf{P}_\nu$. Ta transformacija je usklajena z integracijo.

Izrek B.4.4 nam je dal predpis, ki jedru, usklajenemu s projekcijo p, priredi transformacijo $\nu \mapsto \mathbf{P}_\nu$. Iz implikacije (2) $\Rightarrow$ (1) sledi, da je $p_* \mathbf{P}_\nu = \nu$ (za mero μ na merljivem prostoru $(S, \mathcal{S})$ in merljivo preslikavo $f\colon (S, \mathcal{S}) \to (T, \mathcal{T})$ z $f_* \mu$ označimo mero na $(T, \mathcal{T})$, ki deluje po predpisu $f_* \mu(B) := \mu(f^{-1}(A))$). Transformacija $\nu \mapsto \mathbf{P}_\nu$ je torej desni inverz preslikave p_*.

Izrek B.4.6. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora in naj bo $p \mapsto (T, \mathcal{T}) \to (S, \mathcal{S})$ merljiva preslikava. Dana naj bo transformacija $\nu \mapsto \mathbf{P}_\nu$ iz prostora verjetnostnih mer na $(S, \mathcal{S})$ v prostor verjetnostnih mer na $(T, \mathcal{T})$, ki naj bo desni inverz preslikave p_* in usklajena z integracijo. Definirajmo jedro q iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$ po predpisu $q(x, B) := \mathbf{P}_x(B)$, to je kot v enačbi (B.4.2). Tedaj je jedro q usklajeno s p in za poljubno porazdelitev ν na $(S, \mathcal{S})$ in slučajno spremenljivko $Y\colon (\Omega, \mathcal{F}) \to (T, \mathcal{T})$ s porazdelitvijo $\mathbf{P}_\nu$ velja, da je q prehodna porazdelitev slučajne spremenljivke Y glede na $p(Y)$.*

Dokaz. Ker je $\nu \mapsto \mathbf{P}_\nu$ desni inverz preslikave p_*, je za vsak x tudi $p_* \mathbf{P}_x = \delta_x$. Sledi:

$$q\big(x, p^{-1}(\{x\})\big) = \mathbf{P}_x\big(p^{-1}(\{x\})\big) = (p_* \mathbf{P}_x)(\{x\}) = 1$$

Jedro q je torej res usklajeno s projekcijo p. Dokazali smo že, da jedro q ter porazdelitvi ν in P_ν povezuje enačba (B.4.1). Po točkah (2) in (3) izreka B.4.4 je potem q res prehodna porazdelitev slučajne spremenljivke Y glede na $p(Y)$. ∎

Govoriti o jedrih iz $(S, \mathcal{S})$ v $(T, \mathcal{T})$, usklajenih s projekcijo $p\colon (T, \mathcal{T}) \to (S, \mathcal{S})$, je torej isto kot govoriti o transformacijah $\nu \mapsto \mathbf{P}_\nu$ iz prostora porazdelitev na $(S, \mathcal{S})$ v prostor porazdelitev na $(T, \mathcal{T})$, ki so desni inverzi preslikave p_* in usklajene z integracijo. Tokrat pa ekvivalenca ni račun brez krčmarja, saj se lepo ozraža na slučajnih spremenljivkah.

B.5 Definicija markovskega procesa

Govorili bomo o slučajnih procesih v zveznem času. Slučajni proces bo torej slučajna funkcija iz $[0, \infty)$ v dani merljivi prostor $(S, \mathcal{S})$, ki mu bomo rekli *prostor stanj*. To bo torej slučajna spremenljivka z vrednostmi v $S^{[0,\infty)}$.

Brž ko želimo govoriti o verjetnosti, moramo prostor $S^{[0,\infty)}$ opremiti s primerno σ-algebro. Kar ponuja se nam produktna σ-algebra $\mathcal{S}^{[0,\infty)}$. V tem primeru nam izrek Daniella in Kolmogorova (glej [21], strani 124–127) zagotavlja obstoj lepega števila mer. Vendar pa se izkaže, da tako dobljen merljiv prostor ni preveč uporaben. Če je npr. S topološki prostor in $\mathcal{S}$ Borelova σ-algebra in elemente produkta $S^{[0,\infty)}$ gledamo kot preslikave $[0, \infty) \to S$, je množica vseh zveznih funkcij nemerljiva. Merljive so namreč le množice, ki so odvisne kvečjemu od števno mnogo komponent. Natančneje, če z Y_t označimo koordinatno projekcijo na prostor z indeksom t, so merljive množice natančno kvečjemu števni preseki množic oblike $[Y_t \in A]$, kjer je $A \in \mathcal{S}$. Množica zveznih funkcij pa razen v trivialnem primeru ni take oblike.

Namesto merljivega prostora $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ bomo torej vzeli kak njegov podprostor $(\overline{S}, \overline{\mathcal{S}})$, pri čemer je $\overline{S}$ lahko tudi nemerljiva podmnožica prostora $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$, σ-algebra $\overline{\mathcal{S}}$ pa bo vedno inducirana s σ-algebro $\mathcal{S}^{[0,\infty)}$. To bo torej najmanjša σ-algebra, glede na katero so koordinatne projekcije Y_t merljive. V skladu z zgoraj povedanim definirajmo:

Definicija. Naj bo $(S, \mathcal{S})$ merljiv prostor. *Slučajni proces* je slučajna spremenljivka $X\colon (\Omega, \mathcal{F}) \to (\overline{S}, \overline{\mathcal{S}})$, kjer je $\overline{S} \subset S^{[0,\infty)}$ in $\overline{\mathcal{S}}$ najmanjša σ-algebra na $\overline{S}$, glede na katero so vse koordinatne projekcije Y_t merljive. Za dani $\omega \in \Omega$ je torej $X(\omega) \in \overline{S}$. Elementom $X(\omega)$ bomo rekli trajektorije, prostoru $(\overline{S}, \overline{\mathcal{S}})$ pa *prostor trajektorij*.

Opomba. Z $Y_t\colon \overline{S} \to S$ bomo ves čas označevali koordinatno projekcijo na faktor z indeksom t. Prav tako bomo označili $X_t := Y_t \circ X$.

Navadno nas ne bodo zanimale konkretne slučajne spremenljivke X, ampak le primerne porazdelitve na $(\overline{S}, \overline{\mathcal{S}})$. Težišče naslednjih razdelkov bo torej konstrukcija porazdelitev na tem merljivem prostoru. Ko bomo konstruirali porazdelitev, bo $(Y_t)_{t\in[0,\infty)}$ iskani slučajni proces.

Dostikrat pa nas niti porazdelitve same ne bodo prav dosti zanimale, ampak nas bo bolj zanimal razvoj procesa, če začnemo v dani točki $x \in S$. Zanimala nas bo torej prehodna porazdelitev slučajnega procesa X glede na X_0. Iskali bomo torej jedra iz $(S, \mathcal{S})$ v $(\overline{S}, \overline{\mathcal{S}})$, usklajena s projekcijo Y_0. Takemu jedru bomo rekli *jedro slučajnega procesa*. Pripomnimo naj še, da je po izrekih B.4.4 in B.4.6 govoriti o jedrih slučajnih procesov isto kot govoriti o transformacijah $\nu \mapsto \mathbf{P}_\nu$ iz prostora porazdelitev na $(S, \mathcal{S})$ v prostor porazdelitev na $(\overline{S}, \overline{\mathcal{S}})$, ki so desni inverzi projekcije Y_{0*} in usklajene z integracijo.

Definicija. Če je X slučajni proces in je q prehodna porazdelitev slučajne spremenljivke X glede na X_0, pravimo, da je q *(globalno) jedro* slučajnega procesa X (za razliko od lokalnih jeder, ki bodo prehodne porazdelitve slučajnih spremenljivk X_t glede na X_0 in bodo postale pomembne malo kasneje). Paru (X, q) pravimo *slučajni proces z jedrom* q.

Brž ko imamo dano porazdelitev ν na $(S, \mathcal{S})$, ki ji pravimo *začetna porazdelitev*, in jedro q, ki je jedro slučajnega procesa, lahko torej po izreku B.4.4 konstruiramo slučajni proces z jedrom q.

Zanimali nas bodo *markovski procesi*, to so slučajni procesi, pri katerih je nadaljnji razvoj odvisen le od trenutnega stanja, ne pa tudi od preteklosti, in je tudi neodvisen od trenutka, v katerem gledamo vse skupaj. Opišimo to še formalno.

Definicija. *Levi pomik* je preslikava $\theta_t\colon S^{[0,\infty)} \to S^{[0,\infty)}$, ki deluje po predpisu:

$$\theta_t(x)(s) := x(s+t)$$

Elemente produktnega prostora smo tu pisali kot funkcije iz $[0,\infty)$ v S.

Opomba. *Brž ko je preslikava θ_t dobro definirana kot preslikava $\overline{S} \to \overline{S}$ (t. j. takrat, ko je $\theta_t(\overline{S}) \subset \overline{S}$), je merljiva.* Dovolj je namreč preveriti merljivost množic oblike $\theta_t^{-1}(Y_s^{-1}(A)) = Y_{t+s}^{-1}(A)$, ki je očitna.

Definicija. Slučajni proces X je *markovski*, če za vsak $t \in [0,\infty)$ velja:

(1) $\theta_t(\overline{S}) \subset \overline{S}$.

(2) Slučajna spremenljivka $\theta_t(X)$ je pogojno glede na X_t neodvisna od σ-algebre $\mathcal{F}_t$, ki jo generirajo slučajne spremenljivke X_s, $s \le t$.

Spomnimo se: dogodka A in B sta *pogojno neodvisna* glede na σ-algebro $\mathcal{G}$, če velja:

$$\mathbf{P}(A \cap B \mid \mathcal{G}) = \mathbf{P}(A \mid \mathcal{G})\, \mathbf{P}(B \mid \mathcal{G})$$

σ-algebri $\mathcal{A}$ in $\mathcal{B}$ sta pogojno neodvisni glede na $\mathcal{G}$, če sta poljubna dogodka $A \in \mathcal{A}$ in $B \in \mathcal{B}$ pogojno neodvisna glede na $\mathcal{G}$.

Če pri pogojni neodvisnosti namesto σ-algebre vzamemo slučajno spremenljivko, mislimo s tem σ-algebro, ki jo ta slučajna spremenljivka generira.

Trditev B.5.1. *Če je slučajni proces X markovski, je to tudi $\theta_t(X)$.*

Dokaz. Naj bo $X' := \theta_t(X)$. Treba je preveriti, da je za vsak $s \in [0,\infty)$ slučajna spremenljivka $\theta_s(X') = \theta_{t+s}(X)$ pogojno glede na $X'_s = X_{t+s}$ neodvisna od σ-algebre, ki jo generirajo slučajne spremenljivke $X'_u = X_{t+u}$, $u \le s$. To pa takoj sledi iz dejstva, da je X markovski proces. ∎

V zvezi s pogojno neodvisnostjo je ključen naslednji razmeroma preprost izrek, ki je dokazan v [8] (stran 217, izrek 7.3.1).

Izrek B.5.2. *Naj bo $X\colon (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ slučajna spremenljivka in naj bosta $\mathcal{G}, \mathcal{H}$ σ-algebri. Naslednje trditve so ekvivalentne:*

(1) *X je pogojno glede na $\mathcal{G}$ neodvisna od $\mathcal{H}$.*

(2) *Za vsak $A \in \mathcal{S}$ velja:*

$$\mathbf{P}\big([X \in A] \,\big|\, \sigma(\mathcal{G}, \mathcal{H})\big) = \mathbf{P}\big([X \in A] \,\big|\, \mathcal{G}\big)$$

(3) *Za vsako merljivo funkcijo $f\colon S \to \mathbb{R}$ velja:*

$$\mathbf{E}\big(f(X) \,\big|\, \sigma(\mathcal{G}, \mathcal{H})\big) = \mathbf{E}\big(f(X) \,\big|\, \mathcal{G}\big)$$

■

Opomba. S $\sigma(\mathcal{G},\mathcal{H})$ smo označili najmanjšo σ-algebro, ki vsebuje $\mathcal{G}\cup\mathcal{H}$. Podobno bomo tudi s $\sigma(X)$ označevali najmanjšo σ-algebro na Ω, glede na katero je preslikava $X\colon \Omega \to (S,\mathcal{S})$ merljiva.

Markovske procese torej karakterizira enačba:

$$\mathbf{P}\big([\theta_t(X)\in B]\bigm| \mathcal{F}_t\big) = \mathbf{P}\big([\theta_t(X)\in B]\bigm| X_t\big) \tag{B.5.1}$$

Sledi:

$$\mathbf{E}\big(f(\theta_t(X))\bigm| \mathcal{F}_t\big) = \mathbf{E}\big(f(\theta_t(X))\bigm| X_t\big) \tag{B.5.2}$$

Tu je f zaenkrat indikator poljubne množice $B\in\overline{\mathcal{S}}$. Iz linearnosti in izreka o dominirani konvergenci za pogojno matematično upanje pa sledi, da enakost velja za poljubno funkcijo $f\in\flat(\overline{S},\overline{\mathcal{S}})$.

Kot smo že omenili, nas bodo zanimali le tisti markovski procesi, pri katerih je nadaljnji razvoj neodvisen tudi od trenutka, v katerem gledamo vse skupaj. Razvoj slučajnega procesa pa opisuje jedro, ki je primerna prehodna porazdelitev. Časovno homogenost bo torej naravno definirati za *jedra* slučajnih procesov.

Definicija. Naj bosta $(S,\mathcal{S})$ in $(\overline{S},\overline{\mathcal{S}})$ kot zgoraj in naj bo $\theta_t(\overline{S})\subset\overline{S}$ za vsak $t\in[0,\infty)$. Jedro q iz $(S,\mathcal{S})$ v $\overline{S},\overline{\mathcal{S}})$ je *markovsko*, če je za vsak slučajni proces X z jedrom q in vsak $t\in[0,\infty)$ slučajni proces $\theta_t(X)$ markovski in ima q za jedro.

Časovno homogen markovski proces je tak slučajni proces X, da obstaja markovsko jedro q, ki je jedro slučajnega procesa X, torej prehodna porazdelitev slučajne spremenljivke X glede na X_0.

Opomba. Po trditvi B.5.1 je za to, da je jedro q markovsko, dovolj preveriti, da je vsak slučajni proces X z jedrom q markovski in da je za vsak $t\in[0,\infty)$ slučajna spremenljivka $\theta_t(X)$ slučajni proces z jedrom q.

Naj bo q markovsko jedro in X slučajni proces z jedrom q. Iz dejstva, da je slučajni proces X markovski, enačbe (B.5.2), dejstva, da je tudi $\theta_t(X)$ slučajni proces z jedrom q, in trditve B.4.1 sledi, da za vsako funkcijo $f\in\flat(\overline{S},\overline{\mathcal{S}})$ velja:

$$\mathbf{E}\big(f(\theta_t(X))\bigm| \mathcal{F}_t\big) = \mathbf{E}\big(f(\theta_t(X))\bigm| X_t\big) = \mathbf{E}_{X_t}(f)$$

Slučajna spremenljivka $\mathbf{E}_{X_t}(f)$ je tu definirana kot kompozitum $\phi\circ X_t$, kjer je $\phi(x) := \mathbf{E}_x(f)$. Videli smo že, da je funkcija ϕ merljiva.

Kar smo dobili, je *lastnost Markova*. Le-ta tudi karakterizira markovska jedra.

Trditev B.5.3. *Naj bo q jedro slučajnega procesa in naj z oznakami od prej velja, da je $\theta_t(\overline{S})\subset\overline{S}$ za vsak $t\in[0,\infty)$. Naslednji trditvi sta ekvivalentni:*

(1) *Jedro q je markovsko.*

(2) *Za vsak slučajni proces X z jedrom q velja lastnost Markova:*

$$\mathbf{E}_{X_t}(f) = \mathbf{E}\big(f(\theta_t(X))\bigm| \mathcal{F}_t\big) \tag{B.5.3}$$

DOKAZ. Implikacijo (1) $\Rightarrow$ (2) smo že dokazali. Naj zdaj velja (2). Ker je slučajna spremenljivka $\mathbf{E}_{X_t}(f)$ merljiva glede na $\sigma(X_t)$, velja:

$$\mathbf{E}_{X_t}(f) = \mathbf{E}\big(f(\theta_t(X)) \bigm| X_t\big) = \mathbf{E}\big(f(\theta_t(X)) \bigm| \mathcal{F}_t\big)$$

Iz prvega enačaja in trditve B.4.1 zdaj sledi, da je $\theta_t(X)$ slučajni proces z jedrom q za vsak t, iz drugega enačaja in enačbe (B.5.2) pa sledi, da je slučajni proces X markovski. Iz trditve B.5.1 potem sledi, da je q res markovsko jedro. ∎

Opomba. V literaturi je s pojmom markovski proces dostikrat mišljeno markovsko jedro, definicija pa temelji kar na enačbi (B.5.3). Tako je tudi v [21], kjer se markovsko jedro q opiše kot zbirka porazdelitev $\mathbf{P}_x$. Zveza med markovskim jedrom q in porazdelitvami $\mathbf{P}_x$ je podana s formulo (B.4.2): $q(x, B) = \mathbf{P}_x(B)$. No, definicija v [21] je še nekoliko splošnejša: prostor trajektorij ni nujno podprostor prostora $S^{[0,\infty)}$ in tako tudi Y_t niso nujno koordinatne projekcije. Splošneje so vzete tudi σ-algebre $\mathcal{F}_t$. Naša definicija pa se sklada s tisto iz [10].

B.6 Markovski procesi in operatorske polgrupe

Definirali smo, kaj je markovski proces. Vendar pa zaenkrat razen nekaj trivialnih zgledov, kot so npr. konstante, ne bi mogli konstruirati nobenega markovskega procesa.

Kako sploh konstruirati kako markovsko jedro? To je jedro iz (navadno krotkega) prostora stanj $(S, \mathcal{S})$ v ogromni prostor $(\overline{S}, \overline{\mathcal{S}})$. Izkazalo pa se bo, da je mogoče to markovsko jedro, ki je, kot vemo, prehodna porazdelitev slučajne spremenljivke X glede na X_0, kjer je X slučajni proces z jedrom q, izluščiti že iz *lokalnih* jeder, ki so prehodne porazdelitve slučajnih spremenljivk X_t glede na X_0.

Kako iz globalnega jedra dobiti lokalna, vemo. Danemu jedru q iz $(S, \mathcal{S})$ v $(\overline{S}, \overline{\mathcal{S}})$ priredimo jedra q_t na prostoru stanj $(S, \mathcal{S})$ po predpisu:

$$q_t(x, A) := q(x, [Y_t \in A]) \tag{B.6.1}$$

Iz izreka B.2.3 takoj sledi: če je X slučajni proces z jedrom q, je q_t prehodna porazdelitev slučajne spremenljivke X_t glede na X_0.

Posebno če je jedro q markovsko, jedra q_t ne morejo biti kar koli. Izkaže se, da jih karakterizirata dve osnovni lastnosti. V primeru, ko je $(\overline{S}, \overline{\mathcal{S}}) = (S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$, bosta ti dve lastnosti dovolj: vsaki družini $(q_t)_{t\in[0,\infty)}$, ki ju izpolnjuje, pripada natančno eno markovsko jedro q, za katerega velja enačba (B.6.1). Ta 'produktni' primer bo tudi osnova za naslednji zanimivejši primer, ko bo S dovolj lep topološki prostor z Borelovo σ-algebro $\mathcal{S}$, $(\overline{S}, \overline{\mathcal{S}})$ pa bo prostor vseh funkcij iz $[0, \infty)$ v S, ki so z desne zvezne in imajo limite z leve. To bo izplonjeno tudi pri *Feller–Dynkinovih procesih*), ki bodo za nas še posebej zanimivi. V tem primeru pa bo morala družina $(q_t)_{t\in[0,\infty)}$ izpolnjevati še dodatne pogoje.

Pa si oglejmo ti dve osnovni lastnosti, ki morajo veljati za jedra q_t. Za vsak $x \in S$ pripada markovskemu jedru q porazdelitev $\mathbf{P}_x$. Če merljivi prostor $(\overline{S}, \overline{\mathcal{S}})$ opremimo z verjetnostno mero $\mathbf{P}_x$, postanejo projekcije Y_t slučajne spremenljivke in izrek B.4.4 nam pove, da je $Y := (Y_t)_{t\in[0,\infty)}$ (ki je identiteta na $\overline{S}$) slučajni proces z jedrom q in začetno porazdelitvijo δ_x.

Iz dejstva, da je jedro q_t je prehodna porazdelitev slučajne spremenljivke Y_t glede na Y_0, in enačbe (B.1.1) sledi:

$$q_t(x, A) = \int q_0(y, A)\mathcal{L}(Y_0)(dy) = \mathbf{P}_x[Y_t \in A]$$

Velja torej $\mathcal{L}(Y_t)(A) = q_t(x, A)$. Ker je slučajna spremenljivka Y_0 skoncentrirana v x, velja:

$$q_0(x, A) = I_A(x) \tag{B.6.2}$$

To je prva lastnost, ki karakterizira jedra q_t. Jedro q_0 torej predpisuje: če je q_0 prehodna porazdelitev slučajne spremenljivke U glede na V, je skoraj gotovo $U = V$.

Videli smo, da je $\mathcal{L}(Y_t)(A) = q_t(x, A)$. Torej je tudi $\mathcal{L}(Y_{t+s})(A) = q_{t+s}(x, A)$. Iz dejstva, da je tudi $\theta_t(Y)$ slučajni proces z jedrom q, sledi, da je q_s prehodna porazdelitev slučajne spremenljivke Y_{t+s} glede na Y_t. Od tod pa dobimo:

$$q_{t+s}(x, A) = \mathbf{P}_x[Y_{t+s} \in A] = \int q_s(y, A)\, \mathcal{L}(Y_t)(dy) = \int q_s(y, A)\, q_t(x, dy)$$

Dobljeni enačbi:

$$q_{t+s}(x, A) = \int q_s(y, A)\, q_t(x, dy) \tag{B.6.3}$$

pravimo *enačba Chapmana in Kolmogorova* in to je druga lastnost, ki karakterizira jedra q_t.

Po posledici B.3.3 pa vsakemu jedru q_t pripada natančno določen linearni operator $P_t\colon \flat(S, \mathcal{S}) \to \flat(S, \mathcal{S})$, definiran po predpisu:

$$(P_t f)(x) = \int f(y)\, q(x, dy)$$

Izkazalo se bo namreč, da je za nadaljnjo obravnavo in konstrukcijo družin jeder q_t precej ugodneje gledati njim pripadajoče linearne operatorje.

Opomba. Če je X slučajni proces z jedrom q, ki je markovsko, je torej $P_t f$ regresija slučajne spremenljivke $f(X_t)$ glede na X_0. Z drugimi besedami, velja $(P_t f)(X_0) = E\big(f(X_t) \bigm| X_0\big)$.

Oglejmo si, kako se enačbi (B.6.2) in (B.6.3) odražata na dobljenih operatorjih. Prva enačba očitno pove, da je P_0 kar identiteta. Druga enačba pa najprej pove, da v primeru, ko je $f = I_A$, velja:

$$(P_{t+s} f)(x) = q_{t+s}(x, A) = \int q_s(y, A)\, q_t(x, dy) = \int\!\!\int f(z)\, q_s(y, dz)\, q_t(x, dy) = (P_t P_s f)(x)$$

V primeru, ko je f indikator, je torej $P_{t+s} f = P_t P_s f$. Zaradi linearnosti in monotone konvergence pa to velja tudi za vsak $f \in \flat(S, \mathcal{S})$. Velja torej $P_{t+s} = P_t P_s$.

Definicija. *Operatorska polgrupa* na Banachovem prostoru L je družina $(P_t)_{t \in [0,\infty)}$ omejenih linearnih operatorjev na L, ki ima naslednji lastnosti:

(1) $P_0 = I$.

(2) $P_{t+s} = P_t P_s$.

Vsakemu markovskemu jedru smo torej priredili operatorsko polgrupo na $\flat(S, \mathcal{S})$. Operatorji, ki sestavljajo to polgrupo, so poleg tega še pozitivni, normalizirani in zaprti za monotono konvergenco. Nastane vprašanje, ali je mogoč tudi obrat. Izrek B.3.5 nam vsaki operatorski polgrupi z zgornjimi lastnostmi priredi družino jeder q_t na $(S, \mathcal{S})$. Nastane seveda vprašanje, ali zdaj veljata enačbi (B.6.2) in (B.6.3). Odgovor je po pričakovanju pritrdilen. Prva enačba takoj sledi iz lastnosti (1), lastnost (2) pa nam pove:

$$\begin{aligned} q_{t+s}(x, A) = (P_{t+s} I_A)(x) = (P_t P_s I_A)(x) = \int \int I_A(z)\, q_s(y, dz)\, q_t(x, dy) = \\ = \int q_s(y, A)\, q_t(x, dy) \end{aligned}$$

kar je natančno enačba Chapmana in Kolmogorova. *Govoriti o družinah jeder na $(S, \mathcal{S})$, ki zadoščajo enačbama (B.6.2) in (B.6.3) je torej isto kot govoriti o operatorskih polgrupah na $\flat(S, \mathcal{S})$, katere operatorji so pozitivni, normalizirani in zaprti za monotono konvergenco.* Drugo vprašanje pa je, ali je to tudi isto kot govoriti o markovskih jedrih. V naslednjem razdelku bomo pokazali, da je odgovor pritrdilen v primeru, ko je $(\overline{S}, \overline{\mathcal{S}}) = (S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$.

B.7 Konstrukcija markovskih procesov

Cilj tega razdelka bo pokazati naslednji izrek.

Izrek B.7.1. *Naj bo $(S, \mathcal{S})$ Borelov prostor in naj bo $(q_t)_{t \in [0,\infty)}$ družina jeder na $(S, \mathcal{S})$, ki zadoščajo enačbama (B.6.2) in (B.6.3), torej:*

$$\begin{aligned} q_0(x, A) &= I_A(x) \\ q_{t+s}(x, A) &= \int q_s(y, A)\, q_t(x, dy) \end{aligned}$$

Tedaj obstaja natanko eno markovsko jedro q iz $(S, \mathcal{S})$ v $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$, ki mu pripadajo lokalna jedra q_t, t. j. veljati mora enačba (B.6.1):

$$q_t(x, A) = q\big(x, [Y_t \in A)\big)$$

Z drugimi besedami, da vsak slučajni proces X z jedrom q mora biti q_t prehodna porazdelitev slučajne spremenljivke X_t glede na X_0.

Jedro q bomo konstruirali kot primerno transformacijo $\nu \mapsto \mathbf{P}_\nu$ in se oprli na izrek B.4.6. Porazdelitev $\mathbf{P}_\nu$ bomo najprej konstruirali na množicah:

$$[Y_{t_1} \in A_1, \ldots Y_{t_n} \in A_n]$$

ki jim bomo rekli *cilindrične*. Konstrukcijo bomo izvedli iz posledice B.4.5. Iz nje in trditve B.4.2 zlahka izpeljemo naslednjo trditev.

Trditev B.7.2. *Naj bodo $(S_k, \mathcal{S}_k)$, $k = 0, \dots n$ merljivi prostori. Naj bo ν porazdelitev na $(S_0, \mathcal{S}_0)$ in za $k = 1, \dots n$ naj bo q_k jedro iz $\left(\prod_{i=0}^{k-1} S_i, \prod_{i=1}^{k-1} \mathcal{S}_i\right)$ v $(S_k, \mathcal{S}_k)$. Tedaj obstaja natanko ena porazdelitev $\mathbf{P}_\nu$ na $\left(\prod_{i=0}^{n} S_i, \prod_{i=0}^{n} \mathcal{S}_i\right)$, za katero velja, da sta poljubne slučajne spremenljivke X_i: $(\Omega, \mathcal{F}) \to (S_i, \mathcal{S}_i)$, $i = 0, \dots n$, naslednji trditvi ekvivalentni:*

(1) *Slučajna spremenljivka X_0 ima porazdelitev ν, q_k pa je prehodna porazdelitev slučajne spremenljivke X_k glede na $(X_0, \dots X_{k-1})$ za vsak k.*

(2) *Slučajno zaporedje $(X_0, \dots X_n)$ ima porazdelitev $\mathbf{P}_\nu$.*

Transformacija $\nu \mapsto \mathbf{P}_\nu$, ki pripada jedrom q_k, je usklajena z integracijo. ■

Za nas je seveda zanimiv markovski primer.

Definicija. Zaporedje slučajnih spremenljivk $(X_0, \dots X_n)$ je *markovsko*, če je za vsak $k = 1, \dots n-1$ slučajna spremenljivka X_{k+1} pogojno glede na X_k neodvisna od slučajnega zaporedja $(X_0, \dots X_{k-1})$.

Pri dokazovanju markovske lastnosti konstruiranega slučajnega procesa pa bomo potrebovali malo močnejšo karakterizacijo.

Trditev B.7.3. *Če je zaporedje $(X_0, \dots X_n)$ slučajnih spremenljivk z vrednostmi v Borelovih prostorih markovsko, sta za vsak $k = 0, \dots n$ pogojno glede na X_k neodvisni tudi slučajni zaporedji $(X_0, \dots X_k)$ in $(X_k, \dots X_n)$.*

Dokaz. Za $k = 0$ in $k = n$ je trditev očitna. Za $k = 1, \dots n-1$ pa je dovolj dokazati, da sta pogojno glede na X_k neodvisni slučajni zaporedji $(X_0, \dots X_{k-1})$ in $(X_{k+1}, \dots X_n)$, saj lahko X_k vedno dodamo.

Naj bo torej $k = 1, \dots n-1$ in $k \le i < n$. Tedaj je slučajna spremenljivka X_{i+1} pogojno glede na X_i neodvisna od $(X_0, \dots X_{i-1})$. Iz trditve B.5.2 sledi, da za vsak $A \in \sigma(X_{i+1})$ velja:

$$\mathbf{P}(A \mid \sigma(X_i)) = \mathbf{P}(A \mid \sigma(X_0, \dots X_i))$$

σ-algebri na levi in desni pa lahko zamenjamo tudi z vsako vmesno σ-algebro. Sledi:

$$\mathbf{P}(A \mid \sigma(X_0, \dots X_i)) = \mathbf{P}(A \mid \sigma(X_k, \dots X_i))$$

Po trditvi B.5.2 je potem slučajna spremenljivka X_{i+1} pogojno glede na $\sigma(X_k, \dots X_i)$ neodvisna od $(X_0, \dots X_{k-1})$. Obrnimo to pogojno neodvisnost! Iz trditve B.5.2 zdaj dobimo, da za vsak $B \in \sigma(X_0, \dots X_{k-1})$ velja:

$$\mathbf{P}(B \mid \sigma(X_k, \dots X_i)) = \mathbf{P}(B \mid \sigma(X_k, \dots X_{i+1}))$$

Z indukcijo po i potem sledi, da je tudi:

$$\mathbf{P}(B \mid \sigma(X_k)) = \mathbf{P}(B \mid \sigma(X_k, \dots X_n))$$

od koder spet po trditvi B.5.2 dobimo, da sta slučajni zaporedji $(X_0, \dots X_{k-1})$ in $(X_{k+1}, \dots X_n)$ pogojno glede na X_k res neodvisni. ■

Posledica B.7.4. *Vsako podzaporedje markovskega slučajnega zaporedja je tudi markovsko.*

Za markovski primer dobi trditev B.7.2 naslednjo obliko.

Posledica B.7.5. *Naj bodo $(S_k, \mathcal{S}_k)$, $k = 0, \ldots n$ merljivi prostori. Naj bo ν porazdelitev na $(S_0, \mathcal{S}_0)$ in za $k = 1, \ldots n$ naj bo q_k jedro iz $(S_{k-1}, \mathcal{S}_{k-1})$ v $(S_k, \mathcal{S}_k)$. Tedaj obstaja natanko ena verjetnostna mera $\mathbf{P}_\nu$ na $\left(\prod_{i=0}^n S_i, \prod_{i=0}^n \mathcal{S}_i\right)$, za katero velja, da sta za poljubne slučajne spremenljivke X_i: $(\Omega, \mathcal{F}) \to (S_i, \mathcal{S}_i)$, $i = 0, \ldots n$, naslednji trditvi ekvivalentni:*

(1) *Slučajno zaporedje $(X_0, \ldots X_n)$ je markovsko, X_0 ima porazdelitev ν in za vsak k je q_k prehodna porazdelitev slučajne spremenljivke X_k glede na X_{k-1}.*

(2) *Slučajno zaporedje $(X_0, \ldots X_n)$ ima porazdelitev $\mathbf{P}_\nu$.*

Transformacija $\nu \mapsto \mathbf{P}_\nu$, ki pripada jedrom q_k, je usklajena z integracijo. ■

Dokaz. Definirajmo jedra q'_k iz $\left(\prod_{i=0}^{k-1} S_i, \prod_{i=1}^{k-1} \mathcal{S}_i\right)$ v $(S_k, \mathcal{S}_k)$ po predpisu:

$$q'_k(x_0, \ldots x_{k-1}, A) := q_k(x_{k-1}, A)$$

Iz trditve B.5.2 po kratkem premisleku sledi, da sta za vsako slučajno zaporedje $(X_0, \ldots X_n)$ naslednji trditvi ekvivalentni:

(1) Za vsak k je q'_k prehodna porazdelitev slučajne spremenljivke X_k glede na $(X_0, \ldots X_{k-1})$.

(2) Za vsak k je q_k prehodna porazdelitev slučajne spremenljivke X_k glede na X_{k-1} in slučajno zaporedje $(X_0, \ldots X_n)$ je markovsko.

Potem pa res vse sledi iz trditve B.7.2. ■

Pri dani začetni porazdelitvi in jedrih q_t lahko zdaj konstruiramo porazdelitve slučajnih zaporedij oblike $(X_{t_1}, \ldots X_{t_n})$. Konstruirati pa moramo porazdelitev cele slučajne spremenljivke X. Ključ do tega bo izrek Daniella in Kolmogorova.

Definicija. Naj bo $(S, \mathcal{S})$ merljiv prostor in naj bo T poljubna množica. Družina verjetnostnih mer $(\mu_N)_{N \in F}$ na produktih $(S^N, \mathcal{S}^N)$, kjer je F družina končnih podmnožic množice T, je *usklajena*, če za poljubni množici $M, N \in F$, za kateri je $N \subset M$, velja enakost $\mu_N = p_{MN*}\nu_M$, kjer je p_{MN} naravna projekcija iz S^M na S^N.

Izrek B.7.6 (Daniell, Kolmogorov). *Naj bo $(S, \mathcal{S})$ Borelov prostor, naj bo T poljubna množica in naj bo $(\mu_N)_{N \in F}$ usklajena družina verjetnostnih mer. Tu je F taka družina končnih podmnožic množice T, da je vsaka končna podmnožica v T vsebovana v kakem elementu družine F. Tedaj obstaja natanko ena verjetnostna mera μ na $(S^T, \mathcal{S}^T)$, za katero je $\mu_N = p_{N*}$ za vsak $N \in F$. Tu je p_N naravna projekcija iz S^T na S^N.*

Dokaz bomo opustili, bralec lahko pogleda v [21], strani 124–127. Izrek se najprej dokaže za primer, ko je S kompakten metrični prostor z Borelovo σ-algebro $\mathcal{S}$. Dokaz se skliče na Carathéodoryjev izrek in izrek Tihonova. Iz primera, ko je S kompakten metrični prostor, pa potem takoj sledi splošni primer, saj je vsak Borelov prostor izomorfen merljivi podmnožici kompaktnega metričnega prostora.

Zdaj imamo že pripravljeno vse potrebno, da dokažemo glavni izrek.

Dokaz izreka B.7.1. *Obstoj*: Naj bo F družina vseh končnih podmnožic intervala $[0, \infty)$, ki vsebujejo ničlo. Naj bo ν poljubna porazdelitev na $(S, \mathcal{S})$. Po posledici B.7.5 za vsako množico:

$$N = \{t_0, \dots t_n\} \in F, \quad 0 = t_0 < t_1 < \dots < t_n$$

obstaja taka porazdelitev $\mathbf{P}_\nu^N$, da sta za vsako slučajno spremenljivko $(X_0, \dots X_n)$ z vrednostmi v $(S^{n+1}, \mathcal{S}^{n+1}) \equiv (S^N, \mathcal{S}^N)$ naslednji trditvi ekvivalentni:

(1) $(X_0, \dots X_n)$ ima porazdelitev $\mathbf{P}_\nu^N$.

(2) Slučajno zaporedje $(X_0, \dots X_n)$ je markovsko, X_0 ima porazdelitev ν in za vsak $k = 1, \dots n$ je $q_{t_k - t_{k-1}}$ prehodna porazdelitev slučajne spremenljivke X_k glede na X_{k-1}.

Preverimo, da je družina mer $\mathbf{P}_\nu^N$ usklajena. Brž ko je $N \subset M$, mora torej veljati $\mathbf{P}_\nu^N = p_{MN*}\mathbf{P}_\nu^M$. To pa je dovolj preveriti v primeru, ko se M in N razlikujeta za en sam element. Naj bo torej $N = \{t_0, \dots t_n\}$, $0 = t_0 < t_1 < \dots < t_n$, $M = N \cup \{t\}$ in $t \notin N$. Označimo koordinatne projekcije iz S^M v S takole: za $k = 0, \dots n$ naj bo X_k projekcija na koordinato z indeksom t_k, X pa naj bo projekcija na koordinato z indeksom t. Opremimo $(S^M, \mathcal{S}^M)$ z mero $\mathbf{P}_\nu^M$! Tedaj je dovolj preveriti, da ima slučajno zaporedje $(X_0, \dots X_n)$ porazdelitev $\mathbf{P}_\nu^N$.

Identiteto na $(S^M, \mathcal{S}^M)$ si lahko predstavljamo kot slučajno zaporedje. Le-to je markovsko, torej je markovsko njegovo podzaporedje $(X_0, \dots X_n)$. Prvi 'člen' identitete na S^M je seveda X_0, torej ima X_0 porazdelitev ν. Da preverimo, da ima $(X_0, \dots X_n)$ res porazdelitev $\mathbf{P}_\nu^N$, je treba le še preveriti, da je za vsak $k = 1, \dots n$ jedro q_k prehodna porazdelitev slučajne spremenljivke X_k glede na X_{k-1}.

Slednje je gotovo res, če je $t > t_n$. Naj bo zdaj $t_{r-1} < t < t_r$. Tedaj je $q_{t - t_{r-1}}$ prehodna porazdelitev slučajne spremenljivke X glede na X_{r-1}, $q_{t_r - t}$ pa je prehodna porazdelitev slučajne spremenljivke X_k glede na X. Mi pa moramo preveriti, da je $q_{t_r - t_{r-1}}$ prehodna porazdelitev slučajne spremenljivke X_r glede na X_{r-1}. Iz enačbe Chapmana in Kolmogorova sledi:

$$\begin{aligned}
&\int_A q_{t_r - t_{r-1}}(x, B)\, \mathcal{L}(X_{r-1})(dx_{r-1}) = \\
&\quad = \int_A \int q_{t_r - t}(x, B)\, q_{t - t_{r-1}}(x_{r-1}, dx)\, \mathcal{L}(X_{r-1})(dx_{r-1}) = \\
&\quad = \int \int I_A(x_{k-1})\, q_{t_r - t}(x, B)\, q_{t - t_{r-1}}(x_{r-1}, dx)\, \mathcal{L}(X_{r-1})(dx_{r-1}) = \\
&\quad = \mathbf{E}\big(I_A(X_{k-1})\, q_{t_r - t}(X, B)\big) = \\
&\quad = \int I_A(x_{r-1})\, q_{t_r - t}(x, B)\, \mathcal{L}(X_{r-1}, X)(d(x_{k-1}, x)) = \\
&\quad = \int_{A \times S} q_{t_r - t}(x, B)\, \mathcal{L}(X_{r-1}, X)(d(x_{k-1}, x))
\end{aligned}$$

Tretji enačaj sledi iz dejstva, da je $q_{t-t_{r-1}}$ prehodna porazdelitev slučajne spremenljivke X glede na X_{r-1}, in posledice B.2.2. Vemo pa tudi, da je q_{t_r-t} prehodna porazdelitev slučajne spremenljivke X_r glede na X. Iz dejstva, da je slučajno zaporedje (X_{r-1}, X, X_r) markovsko, in trditve B.5.2 pa sledi, da je funkcija:

$$(x_{r-1}, x, B) \mapsto q_{t_r-t}(x, B)$$

prehodna porazdelitev slučajne spremenljivke X_r glede na (X_{r-1}, X). Potem pa po enačbi (B.1.1) velja:

$$\begin{aligned}\int_A q_{t_r-t_{r-1}}(x, B)\, \mathcal{L}(X_{r-1})(dx_{r-1}) &= \mathbf{P}[(X_{k-1}, X) \in A \times S, X_k \in B] = \\ &= \mathbf{P}[X_{k-1} \in A, X_k \in B]\end{aligned}$$

To pa je tudi enačba (B.1.1), ki pove, da je $q_{t_r-t_{r-1}}$ res prehodna porazdelitev slučajne spremenljivke X_r glede na X_{r-1}, torej je res $\mathbf{P}_\nu^N = p_{MN*}\mathbf{P}_\nu^M$ in družina mer $\mathbf{P}_\nu^N$ je res usklajena.

Po izreku Daniella in Kolmogorova obstaja taka porazdelitev $\mathbf{P}_\nu$ na $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$, da za vsak $N \in F$ velja $\mathbf{P}_\nu^N = p_{N*}\mathbf{P}_\nu$. Opišimo to še malo drugače. Za slučajno spremenljivko X z vrednostmi v $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ kot ponavadi označimo $X_t := Y_t(X)$, kjer so Y_t koordinatne projekcije. Dobili smo: brž ko ima slučajna spremenljivka X porazdelitev $\mathbf{P}_\nu$, ima $(X_{t_0}, \ldots X_{t_n})$ porazdelitev $\mathbf{P}_\nu^N$, kjer je $0 = t_0 < t_1 < \ldots t_n$ in $N = \{t_0, \ldots t_n\}$.

Za vsak N je transformacija $\nu \mapsto \mathbf{P}_\nu^N$ usklajena z integracijo. Od tod sledi, da za vsako množico $B \in \mathcal{S}^{[0,\infty)}$ oblike $[Y_{t_0} \in A_0, \ldots Y_{t_n} \in A_n]$, $0 = t_0 < t_1 < \ldots < t_n$ velja:

$$\mathbf{P}_\nu(B) = \int \mathbf{P}_x(B)\, \nu(dx)$$

saj je $\mathbf{P}_\nu(B) = \mathbf{P}_\nu^N(A_0 \times \ldots \times A_n)$. Množice B zgornje oblike tvorijo π-sistem, ki generira $\mathcal{S}^{[0,\infty)}$. Družina vseh množic, za katere velja enakost, pa je λ-sistem in po Dynkinovi lemi enakost velja za vse množice $B \in \mathcal{S}^{[0,\infty)}$, torej je naša transformacija res usklajena z integracijo. Usklajena pa je tudi s projekcijo Y_0, saj, če prostor $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ opremimo s porazdelitvijo $\mathbf{P}_\nu$, ima Y_0 porazdelitev $\mathbf{P}_\nu^{\{0\}} = \nu$.

Po izreku B.4.6 potem naši transformaciji $\nu \mapsto \mathbf{P}_\nu$ pripada natančno določeno jedro q. Preveriti moramo, da je markovsko. Naj bo X slučajni proces z jedrom q. Trdimo: proces X je markovski. Naj bo ν začetna porazdelitev našega slučajnega procesa. Tedaj ima X porazdelitev $\mathbf{P}_\nu$. Ker ima za vsak $N = \{t_0, \ldots t_n\}$, $0 = t_0 < t_1 < \ldots t_n$, slučajno zaporedje $(X_{t_0}, \ldots X_{t_n})$ porazdelitev $\mathbf{P}_\nu^N$, je markovsko. Od tod pa sledi, da, če je:

$$\begin{aligned} A &= [X_{t_1} \in A_1, \ldots X_{t_m} \in A_m] \\ B &= [X_{s_1} \in B_1, \ldots X_{s_n} \in B_n] \end{aligned}$$

in $0 \le t_1 \le \ldots \le t_m \le t \le s_1 \le \ldots \le s_n$, sta dogodka A in B pogojno glede na X_t neodvisna. Pogojna neodvisnost pa je zaprta za λ-sisteme. Ker dogodki A zgornje oblike tvorijo π-sistem, ki generira $\mathcal{F}_t$, dogodki B pa tvorijo π-sistem, ki generira najmanjšo σ-algebro, v kateri je merljiva slučajna spremenljivka $\theta_t(X)$, iz Dynkinove leme sledi, da je $\theta_t(X)$ pogojno glede na X_t res neodvisna od $\mathcal{F}_t$. Slučajni proces X je torej res markovski.

Preverimo zdaj, da je tudi $\theta_t(X)$ slučajni proces z jedrom q. Privzamemo lahko, da je $t > 0$. Naj bo $t = t_0 < t_1 < \ldots < t_n$. Oglejmo si slučajno zaporedje $(X_{t_0}, \ldots X_{t_n})$. Če mu na začetek pritaknemo še X_0, dobimo, da ima novo slučajno zaporedje porazdelitev $\mathbf{P}_\nu^{N'}$, kjer je $N' = \{0, t_0, \ldots t_n\}$. Od tod pa sledi, da je zaporedje $(X_{t_0}, \ldots X_{t_n})$ markovsko in da je $q_{t_k - t_{k-1}}$ prehodna porazdelitev slučajne spremenljivke X_{t_k} glede na $X_{t_{k-1}}$ za vsak $k = 1, \ldots n$. Če z ν' označimo porazdelitev slučajne spremenljivke X_t, dobimo, da ima potem $(X_{t_0}, \ldots X_{t_n})$ porazdelitev $\mathbf{P}_{\nu'}^N$, kjer je $N = \{t_0 - t, \ldots t_n - t\}$. To pa ne pomeni nič drugega, kot da je $p_{N*}\mu = \mathbf{P}_{\nu'}^N$, kjer je μ porazdelitev slučajnega procesa $\theta_t(X)$. Iz enoličnosti v izreku Daniella in Kolmogorova sledi, da mora biti $\mu = \mathbf{P}_{\nu'}$ in po izreku B.4.6 mora biti potem q prehodna porazdelitev slučajne spremenljivke $\theta_t(X)$ glede na X_t, torej jedro slučajnega procesa $\theta_t(X)$. Dokazali smo, da je jedro q res markovsko.

Preveriti je treba le še enačbo (B.6.1). Za $t = 0$ ta enačba sledi že iz enačbe (B.6.2) in usklajenosti transformacije $\nu \mapsto \mathbf{P}_\nu$ s projekcijo Y_0. Naj bo $t > 0$ in $x \in S$. Opremimo merljiv prostor $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ s porazdelitvijo $\mathbf{P}_x$. Tedaj ima (Y_0, Y_t) porazdelitev $\mathbf{P}_x^{\{0,t\}}$, torej ima Y_0 porazdelitev δ_x, q_t pa je prehodna porazdelitev slučajne spremenljivke Y_t glede naY_0. Sledi:

$$q\big(x, [Y_t \in A]\big) = \mathbf{P}_x[Y_t \in A] = \int q_t(y, A)\, \delta_x(dy) = q_t(x, A)$$

torej (B.6.1) res velja.

Enoličnost: Naj bo q' poljubno markovsko jedro, za katerega velja enačba (B.6.1). Naj bo $x \in S$. Definirajmo porazdelitev μ' po predpisu $\mu'(B) := q(x, B)$ in opremimo merljiv prostor $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ s to porazdelitvijo. Tedaj je $Y := (Y_t)_{t \in [0,\infty)}$ slučajni proces z jedrom q'. Ker je jedro q' markovsko, je q' tudi prehodna porazdelitev slučajne spremenljivke $\theta_t(Y)$ glede na Y_t. Po enačbi (B.6.1) in izreku B.2.3 je tudi q_s prehodna porazdelitev slučajne spremenljivke Y_{t+s} glede na Y_t.

Vzemimo $0 = t_0 < t_1 < \ldots < t_n$ in naj bo $N := \{t_0, \ldots t_n\}$. Slučajno zaporedje $(Y_{t_0}, \ldots Y_{t_n})$ je markovsko, q_{t_k} je prehodna porazdelitev slučajne spremenljivke Y_{t_k} glede na $Y_{t_{k-1}}$ za vsak k in Y_0 ima porazdelitev δ_x. Torej ima $(Y_{t_0}, \ldots Y_{t_n})$ porazdelitev $\mathbf{P}_x^N$. To pa pomeni, da je $p_{N*}\mu' = \mathbf{P}_x^N$. Iz enoličnosti v izreku Daniella in Kolmogorova potem sledi, da je $\mu' = \mathbf{P}_x$ oziroma $q'(x, B) = \mathbf{P}_x(B) = q(x, B)$, torej mora biti $q' = q$. ■

Posledica B.7.7. *Naj bo $(S, \mathcal{S})$ Borelov prostor in naj bo $(P_t)_{t \in [0,\infty)}$ operatorska polgrupa na $\flat(S, \mathcal{S})$. Operatorji naj bodo pozitivni, normalizirani in zaprti za monotono konvergenco. Tedaj obstaja natanko eno jedro q iz $(S, \mathcal{S})$ v $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$, ki je markovsko in mu pripada operatorska polgrupa $(P_t)_{t \in [0,\infty)}$.*

B.8 Feller–Dynkinovi procesi

Kot smo že omenili, markovski procesi v $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ niso pretirano uporabni. Je pa njihova konstrukcija osnova za konstrukcijo zanimivejših markovskih procesov, ki bi si jih zdaj ogledali. Ogledali si bomo *Feller–Dynkinove procese*, pri katerih zahtevamo dvoje:

(1) Trajektorije morajo biti z desne zvezne funkcije z levimi limitami (takim funkcijam bomo rekli *R-funkcije*).

(2) Razvoj procesa mora biti zvezno odvisen od izhodiščne točke.

Kakor hitro se želimo pogovarjati o zveznosti, moramo imeti topologijo. *Zahtevali bomo, da je S kompakten Hausdorffov prostor s števno bazo* (in Borelovo σ-algebro $\mathcal{S}$). Zahteva po kompaktnosti se zdi ostra, vendar se izkaže, da s pojmi, s katerimi smo operirali do sedaj, teorija lepo deluje le za kompaktne prostore. Pri lokalno kompaktnih prostorih si lahko pomagamo s kompaktifikacijo s točko neskončno. Ta točka je potem absorbirajoče stanje markovskega procesa: delec vanjo lahko pride in ko je enkrat tam, tam tudi za vedno obstane. Dejansko se v splošnem lahko zgodi, da delec uide v neskončnost. Če bi se želeli izogniti kompaktifikaciji, bi morali že pri definiciji jedra (glej stran 79) popustiti zahtevo (2): namesto, da je $x \mapsto q(x, B)$ za vsak x verjetnostna mera, bi morali zahtevati le, da je q_x pozitivna mera in da je mera celega prostora največ 1.

Postavimo zdaj definicijo Feller–Dynkinovega procesa na kompaktnem prostoru. Letega pa ne bomo definirali kot marovski proces, temveč kot markovsko *jedro*.

Definicija. *Feller–Dynkinov proces* na kompaktnem Hausdorffovem prostoru je markovsko jedro q iz $(S, \mathcal{S})$ v $(\overline{S}, \overline{\mathcal{S}})$, pri čemer velja:

(1) $\overline{S}$ je prostor vseh R-funkcij iz $[0, \infty)$ v S.

(2) Operatorji P_t operatorske polgrupe, ki pripada jedru q, slikajo zvezne funkcije v zvezne.

Opomba. Včasih bomo s pojmom Feller–Dynkinov proces mislili tudi slučajni proces, ki ima za jedro Feller–Dynkinov proces.

Opomba. Za prostor $\overline{S}$ nismo smeli vzeti produktnega prostora $S^{[0,\infty)}$, saj je množica R-funkcij nemerljiva v $\mathcal{S}^{[0,\infty)}$ in tako zahteve (1) ni mogoče zamenjati z zahtevo: 'trajektorije vsakega slučajnega procesa z jedrom q so skoraj gotovo zvezne'.

Zahteva (2) nam močno olajša delo, saj lahko prostor, na katerem delujejo operatorji P_t, močno zožimo. Velja namreč naslednji izrek.

Izrek B.8.1. *Naj bosta $(S, \mathcal{S})$ in $(T, \mathcal{T})$ merljiva prostora. Prostor T naj bo kompakten Hausdorffov prostor s števno bazo in Borelovo σ-algebro $\mathcal{T}$. Tedaj se da vsaka pozitivna kontrakcija iz $\mathcal{C}(T)$ v $\flat(S, \mathcal{S})$ enolično razširiti do pozitivne kontrakcije iz $\flat(T, \mathcal{T})$ v $\flat(S, \mathcal{S})$, ki je zaprta za monotono konvergenco.*

Dokaz. Naj bo $P\colon \mathcal{C}(T) \to \flat(S, \mathcal{S})$ ustrezni linearni operator. Tedaj je za vsak $x \in S$ preslikava $f \mapsto (Pf)(x)$ pozitivna kontrakcija iz $\mathcal{C}(T)$ v $\mathbb{R}$. Po Rieszovem izreku obstaja taka pozitivna mera μ_x na prostoru $(T, \mathcal{T})$, da za vsako funkcijo $f \in \mathcal{C}(T)$ velja $(Pf)(x) = \int f \, d\mu_x$. Za vsako funkcijo $f \in \flat(T, \mathcal{T})$ bomo torej definirali:

$$(\widehat{P}f)(x) := \int f \, d\mu_x$$

Dobili smo pozitiven linearni operator iz $\flat(T, \mathcal{T})$ v prostor funkcij iz S v $\mathbb{R}$, ki je razširitev operatorja P. Novi operator je očitno tudi zaprt za monotono konvergenco.

Pokažimo, da $\widehat{P}$ slika v $\flat(S, \mathcal{S})$. Ker je P kontrakcija, za vsak $x \in S$ velja $\mu_x(T) = \int 1 \, d\mu_x = (P1)(x) \leq 1$. Od tod pa sledi, da je za vsak $f \in \flat(T, \mathcal{T})$ funkcija $\widehat{P}f$ omejena s supremum normo funkcije f. Dobili smo torej pozitivno kontrakcijo iz $\flat(T, \mathcal{T})$ v prostor omejenih funkcij iz S v $\mathbb{R}$, ki je poleg tega še zaprta za monotono konvergenco.

Naj bo $\mathcal{M}$ družina vseh funkcij iz $\flat(T, \mathcal{T})$, za katere je funkcija $\widehat{P}f$ merljiva. Vemo, da $\mathcal{M}$ vsebuje $\mathcal{C}(T)$. Naj bo zdaj f indikator zaprte množice. Ker je prostor T Hausdorffov in ima števno bazo, je ta zaprta množica presek števne družine odprtih množic. Ker je prostor T normalen, lahko s primerno izbiro Urisonovih funkcij dosežemo, da je f limita padajočega zaporedja zveznih funkcij iz T v $[0, 1]$. Ker je $\widehat{P}$ zaprt za monotono konvergenco (in linearen), je tudi $f \in \mathcal{M}$.

Družina vseh množic $A \in \mathcal{T}$, za katere je $\widehat{P}I_A$ merljiva, je zaradi linearnosti in zaprtosti za monotono konvergenco λ-sistem. Ta λ-sistem pa vsebuje π-sistem vseh zaprtih množic, ki generira $\mathcal{T}$. Po Dynkinovi lemi je potem $I_A \in \mathcal{M}$ za vsak $A \in \mathcal{T}$. Potem pa $\mathcal{T}$ vsebuje vse enostavne funkcije, nadalje vse nenegativne in nazadnje vse funkcije iz $\flat(T, \mathcal{T})$. Operator $\widehat{P}$ torej res slika iz $\flat(T, \mathcal{T})$ v $\flat(S, \mathcal{S})$.

Podobno kot merljivost dokažemo tudi enoličnost. Če je $\widehat{P}'$ še kak operator, ki zadošča pogojem izreka, lahko definiramo množico $\mathcal{N}$ vseh funkcij $f \in \flat(T, \mathcal{T})$, za katere je $\widehat{P}'f = \widehat{P}f$. Podobno kot prej dokažemo, da $\mathcal{N}$ vsebuje najprej indikatorje vseh zaprtih, nato indikatorje vseh merljivih množic in nazadnje vse funkcije iz $\flat(T, \mathcal{T})$. ∎

Operatorje P_t lahko torej res gledamo kot operatorje na $\mathcal{C}(S)$ namesto na $\flat(S, \mathcal{S})$. V skladu s tem bomo s pojmom operatorska polgrupa Feller–Dynkinovega procesa mislili polgrupo operatorjev na $\mathcal{C}(S)$.

Na operatorski polgrupi $(P_t)_{t\in[0,\infty)}$ Feller–Dynkinovega procesa q pa se odraža tudi desna zveznost. Če merljiv prostor $(\overline{S}, \overline{\mathcal{S}})$ opremimo s porazdelitvijo $\mathbf{P}_x$, postane $Y = (Y_t)_{t\in[0,\infty)}$ slučajni proces z začetno porazdelitvijo δ_x in jedrom q. Za vsak f je P_tf regresija slučajne spremenljivke Y_t glede na Y_0, torej je:

$$E(f(Y_t)) = \int P_tf\, d\delta_x = (P_tf)(x)$$

Če je funkcija f zvezna, zaradi desne zveznosti za vsak $\omega \in \overline{S}$ velja:

$$\lim_{t\downarrow 0} f(Y_t(\omega)) = f(Y_0(\omega))$$

Ker je prostor T kompakten, je funkcija f omejena, torej lahko uporabimo izrek o dominirani konvergenci. Sledi:

$$\lim_{t\downarrow 0} E(f(Y_t)) = E(f(Y_0))$$

oziroma:

$$\lim_{t\downarrow 0}(P_tf)(x) = f(x) \tag{B.8.1}$$

Definicija. *Normalizirana Feller–Dynkinova polgrupa* na kompaktnem Hausdorffovem prostoru S s števno bazo je krepko zvezna polgrupa $(P_t)_{t\in[0,\infty)}$ normaliziranih pozitivnih linearnih operatorjev na $\mathcal{C}(S)$. Veljati mora torej:

(1) $P_{t+s} = P_tP_s$ in $P_0 = I$.

(2) $f \geq 0 \Rightarrow P_t f \geq 0$.

(3) $P_t 1 = 1$.

(4) $\lim_{t\downarrow 0} P_t f = f$ za vsak $f \in \mathcal{C}(S)$, in sicer v supremum normi.

Vsakemu Feller–Dynkinovemu procesu na kompaktnem prostoru pripada operatorska polgrupa, ki izpolnjuje zahteve (1)–(3) iz prejšnje definicije, namesto zahteve (4) pa izpolnjuje le šibkejšo zahtevo (B.8.1). Toda izkaže se, da zahteva (B.8.1) že implicira zahtevo (4) iz definicije, kar pove naslednji izrek, ki ga (z dokazom vred) povzemamo iz [20].

Izrek B.8.2. *Naj bo S kompakten Hausdorffov prostor, ki ustreza prvemu aksiomu števnosti (vsaka točka ima števen bazični sistem okolic). Naj bo $(P_t)_{t\in[0,\infty)}$ polgrupa linearnih operatorjev na $\mathcal{C}(S)$ (velja torej $P_0 = I$ in $P_{t+s} = P_t P_s$). Operatorji naj bodo enakomerno omejeni. Obstaja naj torej tak $M > 0$, da je $\|P_t\| \leq M$ za vsak t. Nadalje naj za vsak $f \in \mathcal{C}(S)$ in vsak $x \in S$ velja:*

$$\lim_{t\downarrow 0}(P_t f)(x) = f(x)$$

Tedaj za vsak $f \in \mathcal{C}(S)$ v supremum normi velja:

$$\lim_{t\downarrow 0} P_t f = f$$

Opomba. Pogoj, da so operatorji enakomerno omejeni, se da še oslabiti: dovolj je, da njihove norme eksponentno naraščajo, t. j. $\|P_t\| \leq M e^{\omega t}$. V tem primeru namreč definiramo novo družino linearnih operatorjev P_t' po predpisu $P_t' := e^{-\omega t} P_t$. Ta družina je zdaj enakomerno omejena in tudi polgrupa, zato zanjo velja prejšnji izrek.

Dokaz izreka. Najprej opazimo, da za vsak $t \geq 0$, vsak $f \in \mathcal{C}(S)$ in vsak $x \in S$ velja:

$$\lim_{s\downarrow 0}(P_{t+s} f)(x) = \lim_{s\downarrow 0}(P_s P_t f)(x) = (P_t f)(x)$$

Funkcija $t \mapsto (P_t f)(x)$ je torej z desne zvezna, zato je:

$$(P_t f)(x) = \lim_{n\to\infty}\left(P_{2^{-n}\lceil 2^n t\rceil} f\right)(x)$$

Z $\lceil\cdot\rceil$ smo označili zaokroževanje navzgor. Izraz pod limito je za vsak n in vsak $f \in \mathcal{C}(S)$ merljiva funkcija iz $S \times [0,\infty)$ v $\mathbb{R}$. To pa pomeni, da je tudi funkcija $(x,t) \mapsto (P_t f)(x)$ merljiva. Je pa tudi omejena, zato za vsak $\lambda > 0$ obstaja integral:

$$(R_\lambda f)(x) := \int_0^\infty e^{-\lambda t}(P_t f)(x)\, dt$$

Zgornjemu integralu pravimo *resolventa* (glej trditev B.11.9). Pokažimo, da je za vsak $f \in \mathcal{C}(S)$ tudi $R_\lambda f$ zvezna funkcija. Naj bo $x \in S$. Ker S ustreza prvemu aksiomu števnosti, je dovolj pokazati, da za vsako zaporedje $(x_n)_{n\in\mathbb{N}}$, ki konvergira k x, velja, da $(R_\lambda f)(x_n)$ konvergira k $(R_\lambda f)(x)$. Ker je funkcija $P_t f$ zvezna za vsak t, zaporedje

$(P_t f)(x_n)$ konvergira k $(P_t f)(x)$. Ker je $|(P_t f)(x_n)| \leq M\|f\|$, lahko uporabimo izrek o dominirani konvergenci in dobimo, kar smo iskali. Preslikava R_λ je torej linearni operator na $\mathcal{C}(S)$. Je tudi omejen, saj je $\|R_\lambda\| \leq M\int_0^\infty e^{-\lambda t}\,dt = \frac{M}{\lambda}$. Naj bo D_λ njegova zaloga vrednosti. Pokazali bomo, da je prostor D_λ neodvisen od λ.

Naj bo $t \geq 0$ in $x \in S$. Preslikava $f \mapsto (P_t f)(x)$ je omejen linearni funkcional na $\mathcal{C}(S)$. Po Rieszovem izreku (glej [23], stran 130) obstaja taka omejena realna mera μ na S, da velja $(P_t f)(x) = \int f\,d\mu$. Iz Fubinijevega izreka sledi:

$$\begin{aligned}
(P_t R_\lambda f)(x) &= \int R_\lambda f\,d\mu = \int\int_0^\infty e^{-\lambda s}(P_s f)(y)\,ds\,\mu(dy) = \\
&= \int_0^\infty e^{-\lambda s}\int (P_s f)(y)\,\mu(dy)\,ds = \int_0^\infty e^{-\lambda s}(P_t P_s f)(x)\,ds = \\
&= \int_0^\infty e^{-\lambda s}(P_{t+s} f)(x)\,ds = e^{\lambda t}\int_t^\infty e^{-\lambda s}(P_s f)(x)\,ds
\end{aligned} \tag{B.8.2}$$

Spet iz Fubinijevega izreka sledi:

$$\begin{aligned}
(R_\lambda R_\mu f)(x) &= \int_0^\infty e^{-\lambda t}(P_t R_\mu f)(x)\,dt = \int_0^\infty e^{-\lambda t}e^{\mu t}\int_t^\infty e^{-\mu s}(P_s f)(x)\,ds\,dt = \\
&= \int_0^\infty e^{-\mu s}(P_s f)(x)\int_0^s e^{(\mu-\lambda)t}\,dt = \int_0^\infty e^{-\mu s}(P_s f)(x)\frac{e^{(\mu-\lambda)s}-1}{\mu-\lambda}\,ds = \\
&= \frac{1}{\mu-\lambda}\int_0^\infty \bigl(e^{-\lambda u} - e^{-\mu u}\bigr)(P_u f)(x)\,dx = \frac{1}{\mu-\lambda}\bigl((R_\lambda f)(x) - (R_\mu f)(x)\bigr)
\end{aligned}$$

Dobili smo torej:

$$R_\lambda R_\mu = \frac{1}{\mu-\lambda}(R_\lambda - R_\mu)$$

Naj bo zdaj $\lambda, \mu > 0$ in $f \in D_\mu$. Obstaja torej tak g, da je:

$$f = R_\mu g = R_\lambda g + (\lambda-\mu)R_\lambda R_\mu g$$

Dokazali smo, da je $D_\mu \subset D_\lambda$. Če zamenjamo λ in μ, dobimo še obratno inkluzijo, torej je $D_\lambda = D_\mu$. Prostor D_λ je torej res neodvisen od λ in ga zato označimo kar z D.

Pokažimo, da je D gost v $\mathcal{C}(S)$. Pa recimo, da ni tako. Tedaj obstaja $f \in \mathcal{C}_0(S) \setminus \overline{D}$. Naj bo D' linearna lupina prostora $\overline{D}$ in elementa f. Ker je $\overline{D}$ zaprt v D', obstaja omejen linearni funkcional φ na D', ki je na $\overline{D}$ enak 0 in za katerega je $\varphi(f) = 1$. Po Hahn–Banachovem izreku se da ta funkcional razširiti do omejenega linearnega funkcionala Φ na celem prostoru $\mathcal{C}(S)$. Spet po Rieszovem izreku pa obstaja taka omejena realna mera μ na S, da je $\Phi(g) = \int g\,d\mu$ za vsak $g \in \mathcal{C}(S)$.

Naj bo $x \in S$. Iz izreka o dominirani konvergenci sledi:

$$\begin{aligned}
\lim_{\lambda\to\infty}\lambda(R_\lambda f)(x) &= \lim_{\lambda\to\infty}\lambda\int_0^\infty e^{-\lambda t}(P_t f)(x)\,dt = \lim_{\lambda\to\infty}\int_0^\infty e^{-s}(P_{s/\lambda} f)(x)\,ds = \\
&= \int_0^\infty e^{-s} f(x)\,ds = f(x)
\end{aligned}$$

Ker je $\|R_\lambda\| \le \frac{M}{\lambda}$, je tudi $|\lambda(R_\lambda f)(x)| \le \frac{M}{\lambda}\|f\|$. Spet iz izreka o dominirani konvergenci sledi:

$$1 = \varphi(g) = \int g\, d\mu = \lim_{\lambda\to\infty} \int \lambda\, R_\lambda f\, d\mu = \lim_{\lambda\to\infty} \varphi(\lambda\, R_\lambda f) = 0$$

saj je $\lambda\, R_\lambda f \in D$. To pa je protislovje, torej mora biti prostor D res gost v $\mathcal{C}(S)$.

Spomnimo se zdaj spet na (B.8.2):

$$(P_t R_\lambda f)(x) = e^{\lambda t} \int_t^\infty e^{-\lambda s} (P_s f)(x)\, ds = e^{\lambda t} (R_\lambda f)(x) - e^{\lambda t} \int_0^t e^{-\lambda s} (P_s f)(x)\, ds$$

Sledi:

$$|(P_t R_\lambda f)(x) - (R_\lambda f)(x)| \le \big(e^{\lambda t} - 1\big) |R_\lambda f)(x)| + M e^{\lambda t} \int_0^t e^{-\lambda s}\, ds = \big(e^{\lambda t - 1}\big)\big(|R_\lambda f)(x)| + M\big)$$

Izberimo λ. Izraz na desni strani gre očitno proti 0, torej je:

$$\lim_{t\downarrow 0} \|P_t R_\lambda f - R_\lambda f\| = 0$$

Za vsak $g \in D$ je torej $\lim_{t\downarrow} \|P_t g - g\| = 0$. Naj bo zdaj $\varepsilon > 0$ in $f \in \mathcal{C}(S)$. Ker je podprostor D gost, obstaja tak $g \in D$, da je $\|f - g\| \le \frac{\varepsilon}{2(1+M)}$. Obstaja tudi tak $\delta > 0$, da za vsak $t \in [0, \delta)$ velja $\|P_t g - g\| < \frac{\varepsilon}{2}$. Sledi:

$$\begin{aligned} \|P_t f - f\| &= \|P_t f - P_t g + P_t g - g + g - f\| \le \|P_t\|\, \|f - g\| + \|P_t g - g\| + \|g - f\| < \\ &< M \frac{\varepsilon}{2(M+1)} + \frac{\varepsilon}{2} + \frac{\varepsilon}{2(M+1)} = \varepsilon \end{aligned}$$

Torej je tudi $\lim_{t\downarrow 0} \|P_t f - f\| = 0$. To pa smo hoteli dokazati. ∎

Posledica B.8.3. *Vsakemu Feller–Dynkinovemu procesu na kompaktnem prostoru pripada normalizirana Feller–Dynkinova polgrupa.*

Dostikrat bomo imeli opravka s primerom, ko je prostor S le lokalno kompakten. V tem primeru ga lahko kompaktificiramo s točko ∂. Seveda zahtevamo še, da je S Hausdorffov in da ima števno bazo topologije. Tedaj oboje velja tudi za $\widehat{S}$. Žal se v teoriji ne da izogniti dejstvu, da lahko slučajni proces uide v ∂, lahko pa dosežemo, da je stanje ∂ absorbirajoče. V naslednjih vrsticah bomo karakterizirali, kako se mora obnašati normalizirana Feller–Dynkinova polgrupa na $\widehat{S}$, če želimo, da je stanje ∂ absorbirajoče.

Kdaj je torej dano stanje ∂ absorbirajoče stanje markovskega jedra q? Takrat, ko velja:

$$\mathbf{P}_\partial[Y_t = \partial \text{ za vsak } t \in [0, \infty)] = 1$$

Če je jedro q Feller–Dynkinov proces, lahko uporabimo desno zveznost. V tem primeru je dovolj zahtevati:

$$\mathbf{P}_\partial[Y_t = \partial \text{ za vsak } t \in [0, \infty) \cap \mathbb{Q}] = 1$$

To pa je ekvivalentno zahtevi, da za vsak $t \in [0, \infty) \cap \mathbb{Q}$ velja $\mathbf{P}_\partial[Y_t = \partial] = 1$. Spet zaradi desne zveznosti je to ekvivalentno zahtevi, da to velja za vsak $t \in [0, \infty)$. Na operatorski polgrupi se to odraža tako, da je $(P_t f)(\partial) = f(\partial)$ za vsak t.

Prostor $\mathcal{C}(\widehat{S})$ lahko zapišemo kot direktno vsoto prostora funkcij, ki so v ∂ enake 0, in prostora konstantnih funkcij. Na prostoru konstantnih funkcij vemo, kako se morajo obnašati operatorji P_t: pustiti jih morajo pri miru. Prvi prostor pa lahko identificiramo s $\mathcal{C}_0(S)$ (v primeru, ko je S že kompakten, je $\mathcal{C}_0(S) = \mathcal{C}_(S))$. Od prej vemo: če je stanje ∂ absorbirajoče in je $f(\partial) = 0$, mora biti tudi $(P_t f)(\partial) = 0$. Operatorji P_t morajo torej $\mathcal{C}_0(S)$ slikati v $\mathcal{C}_0(S)$.

Naj bo torej P pozitiven linearni operator iz $\mathcal{C}_0(S)$ v $\mathcal{C}_0(S)$. Razširimo ta operator do linearnega operatorja $\widehat{P}$: $\mathcal{C}(\widehat{S}) \to \mathcal{C}(\widehat{S})$ tako, da definiramo $\widehat{P}1 := 1$. Definirajmo torej:

$$\widehat{P}f := P(f - f(\partial)) + \partial \tag{B.8.3}$$

kjer $\mathcal{C}_0(S)$ identificiramo s tistimi funkcijami iz $\mathcal{C}(\widehat{S})$, ki so v ∂ enake 0. Novi operator je linearen in normaliziran, ni pa nujno pozitiven.

Kaj mora veljati za operator P, če želimo, da je operator $\widehat{P}$ tudi pozitiven? Jasno je, da mora biti P pozitiven. Dokazali pa smo, da je vsak normaliziran pozitiven operator kontrakcija. Torej je $\widehat{P}$ kontrakcija, potem pa mora biti to tudi P. Velja tudi obratno. Naj bo P pozitivna kontrakcija. Vsako funkcijo iz $\mathcal{C}(\widehat{S})$ lahko zapišemo kot vsoto $f + a$, kjer je $f \in \mathcal{C}_0(S)$, a pa je konstanta. Razcepimo $f = f_+ - f_-$. Naj bo $f + a \geq 0$. Tedaj je očitno $a \geq \|f_-\|$, potem pa velja:

$$\widehat{P}(f + a) = Pf_+ - Pf_- + a \geq Pf_+ - Pf_- + \|f_-\| \geq Pf_+ - Pf_- + \|Pf_-\| \geq 0$$

torej je operator $\widehat{P}$ res pozitiven. Sklep: *operator $\widehat{P}$ je pozitiven natanko tedaj, ko je P pozitivna kontrakcija.*

Definicija. *Feller–Dynkinova polgrupa* na lokalno kompaktnem Hausdorffovem prostoru S s števno bazo je krepko zvezna polgrupa $(P_t)_{t\in[0,\infty)}$ pozitivnih kontrakcij na $\mathcal{C}_0(S)$. Veljati mora torej:

(1) $P_{t+s} = P_t P_s$ in $P_0 = I$.

(2) $f \geq 0 \Rightarrow P_t f \geq 0$.

(3) $\|P_t f\| \leq \|f\|$.

(4) $\lim_{t\downarrow 0} P_t f = f$ za vsak $f \in \mathcal{C}_0(S)$, in sicer v supremum normi.

Opomba. Vsaki Feller–Dynkinovi polgrupi $(P_t)_{t\in[0,\infty)}$ na S pripada normalizirana Feller–Dynkinova polgrupa $(\widehat{P}_t)_{t\in[0,\infty)}$ na $\widehat{S}$, ki jo dobimo po formuli (B.8.3).

Tudi izrek B.8.2 se da posplošiti na lokalno kompaktne prostore.

Izrek B.8.4. *Naj bo S lokalno kompakten Hausdorffov prostor s števno bazo topologije. Naj bo $(P_t)_{t\in[0,\infty)}$ polgrupa linearnih operatorjev na $\mathcal{C}_0(S)$ (velja torej $P_0 = I$ in $P_{t+s} = P_t P_s$). Operatorji naj bodo enakomerno omejeni. Obstaja naj torej tak $M > 0$, da je $\|P_t\| \leq M$ za vsak t. Nadalje naj za vsak $f \in \mathcal{C}_0(S)$ in vsak $x \in S$ velja:*

$$\lim_{t\downarrow 0}(P_t f)(x) = f(x)$$

Tedaj za vsak $f \in \mathcal{C}_0(S)$ v supremum normi velja:

$$\lim_{t \downarrow 0} P_t f = f$$

DOKAZ. Ker ima S števno bazo topologije, kompaktifikacija $\widehat{S}$ ustreza prvemu aksiomu števnosti. Definirajmo operatorje $\widehat{P}_t$ na $\widehat{S}$ po formuli (B.8.3). Le-ti tudi tvorijo polgrupo in če je $\|P_t\| \leq M$, je $\|\widehat{P}_t\| \leq 3M$. Nova operatorska polgrupa torej izpolnjuje pogoje izreka B.8.2, iz katerega potem sledi, kar smo hoteli dokazati. ∎

Opomba. Feller–Dynkinova polgrupa je poseben primer krepko zvezne kontrakcijske operatorske polgrupe, ki je definirana na strani 112.

Zdaj, ko smo tako lepo definirali Feller–Dynkinove operatorske polgrupe, se človek seveda vpraša, ali vsaka taka polgrupa določa Feller–Dynkinov proces. Odgovor je pritrdilen, saj velja naslednji izrek.

Izrek B.8.5. *Naj bo $(P_t)_{t \in [0,\infty)}$ normalizirana Feller–Dynkinova polgrupa na S. Tedaj obstaja natanko en Feller–Dynkinov proces, ki mu pripada polgrupa $(P_t)_{t \in [0,\infty)}$.*

SKICA DOKAZA. Po izreku B.7.1 obstaja markovsko jedro q iz $(S, \mathcal{S})$ v $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$, ki mu pripada polgrupa $(P_t)_{t \in [0,\infty)}$ (razširjena na $\flat(S, \mathcal{S})$ po izreku B.8.1). Žal jedro q še ni Feller–Dynkinov proces, saj ni definirano na pravem prostoru. Izkazalo pa se bo, da lahko markovsko jedro q dobimo tako, da z inkluzijo preslikamo neko markovsko jedro iz $(S, \mathcal{S})$ v prostor R-funkcij iz $[0, \infty)$ v S. To jedro bo iskani Feller–Dynkinov proces. Našli pa ga bomo tako, da bomo malo obdelali jedro q.

Elemente prostora $S^{[0,\infty)}$ lahko gledamo kot preslikave. Definirajmo preslikavo $F \colon D \to \overline{S}$ po predpisu:

$$F(x)(t) := \lim_{\substack{q \downarrow t \\ q \in \mathbb{Q}}} x(q)$$

Tu je D množica vseh točk x, za katere zgornja limita obstaja za vsak t in nam definira R-funkcijo. Prostor $\overline{S}$ pa je množica vseh R-funkcij iz $[0, \infty)$ v S.

Markovskemu jedru q pripada transformacija $\nu \mapsto \mathbf{P}_\nu$. S pomočjo martingalov (glej [21], stran 245) se da dokazati, da je množica D merljiva v $\mathcal{S}^{[0,\infty)}$ in da za poljubno porazdelitev ν velja $\mathbf{P}_\nu(D) = 1$. Definirajmo $\mathbf{P}'_\nu := F_*\big(\mathbf{P}_\nu\big|_D\big)$. Tako smo definirali transformacijo $\nu \mapsto \mathbf{P}'_\nu$. Brez težav preverimo, da je usklajena z integracijo in projekcijo Y_0, torej nam določa neko jedro iz $(S, \mathcal{S})$ v $(\overline{S}, \overline{\mathcal{S}})$.

Definirajmo $X_t := Y_t(F)$. Da se dokazati, da glede na vsako mero $\mathbf{P}_\nu$ (zoženo na D) skoraj gotovo velja $X_t = Y_t$. Od tod sledi, da tudi skoraj gotovo velja $(X_{t_1}, \ldots X_{t_n}) = (Y_{t_1}, \ldots Y_{t_n})$. To pa ne pomeni nič drugega, kot da je $p_{N*} i_* \mathbf{P}'_\nu = p_{N*} \mathbf{P}_*$, kjer je i inkluzija iz $\overline{S}$ v $S^{[0,\infty)}$, $N = \{t_1, \ldots t_n\}$ in p_N projekcija z $S^{[0,\infty)}$ na S^N. Iz enoličnosti v izreku Daniella in Kolmogorova potem sledi, da je $\mathbf{P}_\nu = i_* \mathbf{P}'_\nu$. To pa pomeni, da operatorska polgrupa $(P_t)_{t \in [0,\infty)}$ ne pripada le jedru q, ampak tudi jedru q'. V nasprotju s q pa je q' Feller–Dynkinov proces. ∎

Posledica B.8.6. *Naj bo $(P_t)_{t \in [0,\infty)}$ Feller–Dynkinova polgrupa na S. Tedaj obstaja natanko en Feller–Dynkinov proces na kompaktifikaciji $\widehat{S}$ prostora S, ki mu pripada polgrupa $(P_t)_{t \in [0,\infty)}$, naravno razširjena na $\widehat{S}$.*

B.9 Poissonov proces

Zdaj smo pri konstrukciji markovskih procesov že toliko napredovali, da lahko naredimo kak konkreten zgled. Prvi tak zgled bo *Poissonov proces*, ki simulira občasno pojavljanje med seboj neodvisnih dogodkov, npr. telefonskih klicev, impulzov pri idealnem Geigerjevem števcu, rojstev, smrti, nesreč itd. Proces bomo podali tako, da bo X_t število naših dogodkov, ki so se zgodili *od časa* 0 *do vključno časa* t, pri čemer trenutek 0 ni vključen, trenutek t pa je vključen. Trenutek t mora biti vključen, če želimo desno zveznost. Prostor stanj S bo torej kar množica $\mathbb{N}_0$ naravnih števil z ničlo.

Kaj bi pričakovali od jedra q našega procesa X? Izraz $q(n, A)$ pomeni tole: če se je že zgodilo n dogodkov, kolikšna je verjetnost, da bo nadaljnji potek procesa vključen v dogodku A? Dogodek A pa je merljiva množica R-funkcij iz $[0, \infty)$ v $\mathbb{N}_0$. Verjetnost, da bo nadaljnji potek vključen v dogodku A, če se je že zgodilo n dogodkov, bo morala biti enaka verjetnosti, da bo nadaljnji potek vključen v dogodku $A - n$, če se še ni zgodil noben dogodek. Veljati bo torej moralo:

$$q(n, A) = q(0, A - n) \tag{B.9.1}$$

Ta enačba je poseben primer enačbe, ki določa *levo invariantne* slučajne procese.

Definicija. Naj bo $(G, \mathcal{G})$ merljiva grupa z enoto e. Grupa G levo in desno deluje na prostor $(G^{[0,\infty)}, \mathcal{G}^{[0,\infty)})$. Naj bo $\overline{G} \subset G^{[0,\infty)}$ invariantna za levo delovanje in naj bo $\overline{\mathcal{G}}$ najmanjša σ-algebra na $\overline{G}$, glede na katero so vse koordinatne projekcije merljive. Očitno je potem tudi $\overline{\mathcal{G}}$ invariantna za levo delovanje. Jedro q iz $(G, \mathcal{G})$ v $(\overline{G}, \overline{\mathcal{G}})$ je *levo invariantno*, če za vsak $x \in G$ in vsak $A \in \mathcal{G}$ velja:

$$q(x, xA) = q(e, A)$$

Trditev B.9.1. *Naj bo q levo invariantno markovsko jedro na merljivi grupi G. Naj bo X slučajni proces z jedrom q. Tedaj velja:*

(1) *$X_t^{-1}\theta_t(X)$ je slučajni proces z začetno porazdelitvijo δ_e in jedrom q in je neodvisen od σ-algebre $\mathcal{F}_t := \sigma(\{X_s \mid 0 \leq s \leq t\})$.*

(2) *Če je U slučajna spremenljivka z vrednostmi v G in porazdelitvijo ν in je X od nje neodvisen slučajni proces z začetno porazdelitvijo δ_e in jedrom q, je UX slučajni proces z začetno porazdelitvijo ν in jedrom q.*

Dokaz. *(1)*: Ker je jedro q markovsko, je tudi $\theta_t(X)$ slučajni proces z jedrom q, torej je q prehodna porazdelitev slučajne spremenljivke $\theta_t(X)$ glede na X_t. Po izreku B.2.3 je potem funkcija q', definirana po predpisu:

$$q'(x, A) := q(x, xA) = q(e, A)$$

prehodna porazdelitev slučajne spremenljivke $X_t^{-1}\theta_t(X)$ glede na X_t. Toda ta funkcija je neodvisna od x, torej mora biti $X_t^{-1}\theta_t(X)$ neodvisna od X_t. Zaradi tega in pogojne neodvisnosti pa potem po izreku B.5.2 za vsak $A \in \sigma(X_t^{-1}\theta_t(X))$ velja $\mathbf{P}(A) = \mathbf{P}(A|X_t) = \mathbf{P}(A \mid \mathcal{F}_t)$, torej je $X_t^{-1}\theta_t(X)$ res neodvisna od X_t.

Iz formule (B.1.1) zdaj sledi, da je funkcija μ, definirana po predpisu $\mu(A) := q(e, A)$, porazdelitev slučajne spremenljivke $X_t^{-1}\theta_t(X)$. Dokazati moramo le še, da je q prehodna porazdelitev te slučajne spremenljivke glede na $Y_0(X_t^{-1}\theta_t(X)) = X_t^{-1}Y_0(\theta_t(X)) = X_t^{-1}X_t = e$. To je torej kar konstanta, zato je spet po formuli (B.1.1) dovolj preveriti, da je $q(e, A) = \mathbf{P}[X_t^{-1}\theta_t(X) \in A]$. To pa smo pravkar dokazali.

(2): Ker ima slučajni proces X začetno porazdelitev δ_e, ima po formuli (B.1.1) slučajna spremenljivka X porazdelitev μ. Zaradi neodvisnosti je potem funkcija q', definirana po predpisu:

$$q'(x, A) := q(e, A)$$

prehodna porazdelitev te slučajne spremenljivke glede na U. Po izreku B.2.3 je potem prehodna porazdelitev slučajne spremenljivke UX glede na U podana po predpisu:

$$(x, A) \mapsto q'(x, x^{-1}A) = q(e, x^{-1}A) = q(x, A)$$

Torej je q prehodna porazdelitev slučajne spremenljivke UX glede na U. Ker je $Y_0(UX) = UY_0(X) = UX_0 = Ue = U$, je UX res slučajni proces z začetno porazdelitvijo ν in jedrom q. ■

Posledica B.9.2. *Naj bo X slučajni proces z jedrom q, ki naj bo tako kot prej. Tedaj so za poljubne $0 \le t_0 \le t_1 \le \ldots$ slučajne spremenljivke $X_{t_0}^{-1}X_{t_1}$, $X_{t_1}^{-1}X_{t_2}$, $\ldots$ neodvisne. Poleg tega je za $0 \le t \le s$ porazdelitev slučajne spremenljivke $X_t^{-1}X_s$ odvisna le od razlike $s - t$, saj velja $\mathbf{P}[X_t^{-1}X_s \in A] = q_{s-t}(A)$.*

Vrnimo se zdaj k našemu procesu X, pri katerem X_t pomeni število dogodkov, ki so se zgodili do časa t, torej je skoraj gotovo $X_0 = 0$. Kako bi uganili njegovo porazdelitev? Smiselno je privzeti, da se v majhnem časovnem intervalu ne zgodi prav verjetno več kot en dogodek. Razdelimo torej interval od 0 do t na n delov in si oglejmo slučajne spremenljivke $X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}$. Te slučajne spremenljivke so torej približno Bernoullijeve in po prejšnji posledici so neodvisne in enako porazdeljene, X pa je njihova vsota.

Nastane vprašanje, kako postaviti parametre naših Bernoullijevih slučajnih spremenljivk. Definirajmo *pogostnost* pojavljanja dogodkov kot $\lambda := E(X_1)$. Dokazali smo, da je za $0 \le t \le s$ porazdelitev slučajne spremenljivke $X_s - X_t$ odvisna le od razlike $X_s - X_t$. Od tod takoj dobimo, da za poljubno racionalno število t velja $E(X_t) = t\lambda$. Iz desne zveznosti in izreka o dominirani konvergenci pa dobimo, da to velja tudi za vsak realen t. Od tod dobimo, da je slučajno spremenljivko $X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}$ smiselno aproksimirati z Bernoullijevo slučajno spremenljivko, porazdeljeno po $\mathrm{Be}(\frac{\lambda t}{n})$ (za dovolj velik n bo parameter gotovo pod 1). Ker morajo biti slučajne spremenljivke $X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}$ neodvisne, je to smiselno privzeti tudi za njihove aproksimacije. Aproksimacija slučajne spremenljivke X_t je torej porazdeljena po binomski porazdelitvi $\mathrm{Bin}(n, \frac{\lambda t}{n})$. Ko pošljemo n proti neskončno, dobimo, da je smiselno postaviti, da je X_t porazdeljena po Poissonovi porazdelitvi $\mathrm{Po}(\lambda t)$.

Postavili bomo torej:

$$q_t(n, A) := \mathrm{Po}(\lambda t)(A - n)$$

Očitno so q_t jedra. Preveriti je treba le še, da zadoščajo enačbi Chapmana in Kolmogorova. To lahko sicer preverimo neposredno in imamo dobro vajo iz seštevanja vrst. Vendar pa smo enačbo Chapmana in Kolmogorova v resnici že dokazali, le še prav je treba interpretirati zgornje ugotovitve. Naj bo namreč $n \in S$. Vzemimo neodvisni slučajni spremenljivki

Y in Z in naj bo $Y - n$ porazdeljena po $\mathrm{Po}(\lambda t)$, Z pa po $\mathrm{Po}(\lambda s)$. Prehodna porazdelitev slučajne spremenljivke Y glede na Y je potem kar funkcija $(m, A) \mapsto \mathrm{Po}(\lambda s)(A)$. Po izreku B.2.3 je potem q_s prehodna porazdelitev slučajne spremenljivke $Y + Z$ glede na Y. Porazdelitev slednje slučajne spremenljivke pa lahko opišemo kot $\mathbf{P}[Y \in A] = q_t(n, A)$. Ker je vsota neodvisnih Poissonovih slučajnih spremenljivk spet porazdeljena po Poissonu (z vsoto parametrov), velja:

$$q_{t+s}(n, A) = \mathbf{P}[Y + Z \in A] = \int q_s(m, A)\, \mathcal{L}(Y)(dm) = \int q_s(m, A)\, q_t(n, dm)$$

To pa je ravno enačba Chapmana in Kolmogorova. Po izreku B.7.1 potem jedra q_t določajo markovsko jedro. Vendar pa to še ni tisto, kar bi želeli, saj želimo imeti Feller–Dynkinov proces. Za ta namen si moramo ogledati operatorje P_t, ki pripadajo jedrom q_t. Kot vemo, so to pozitivni linearni operatorji na $\flat(S, \mathcal{S})$, ki tvorijo operatorsko polgrupo in delujejo po predpisu:

$$(P_t f)(n) := \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} f(n + k)$$

Ni težko preveriti, da ti operatorji slikajo $\mathcal{C}_0(\mathbb{N}_0)$ spet v $\mathcal{C}_0(\mathbb{N}_0)$ (zvezne so tako ali tako vse funkcije) in da za vsak n velja $\lim_{t \downarrow 0}(P_t f)(n) = f(n)$. To pa pomeni, da ti operatorji tvorijo Feller–Dynkinovo polgrupo. Po posledici B.8.6 obstaja natanko en Feller–Dynkinov proces $\widehat{q}$ na $\widehat{S} := \mathbb{N}_0 \cup \{\infty\}$, ki mu pripada polgrupa operatorjev $\widehat{P}_t$, ki so ustrezne razširitve operatorjev P_t.

Konstruirali smo Feller–Dynkinov proces na $\widehat{S} := \mathbb{N}_0 \cup \{\infty\}$, mi bi pa želeli imeti proces le na $\mathbb{N}_0$. Na sploh ni lahko preveriti, da proces ne uide v neskončnost. Toda tu je to očitno, saj so operatorji P_t dobljeni iz jeder na S, ne pa $\widehat{S}$. Operatorje $\widehat{P}_t$ seveda lahko razširimo tudi na $\flat(\widehat{S}, \widehat{\mathcal{S}})$. Definirajmo funkcije $f_n := I_{\{0,\ldots n\}}$. Te funkcije lahko gledamo v prostoru $\mathcal{C}_0(S)$, pa tudi v $\mathcal{C}(\widehat{S})$. Za vsak $n \in S$ velja $(\widehat{P}_t f_n)(n) = (P_t f_n)(n)$. Ko pošljemo n proti neskončno, iz zaprtosti za monotono konvergenco dobimo, da za vsak $n \in S$ velja $(\widehat{P}_t I_S)(n) = (P_t 1)(n) = 1$. To pa pomeni, da za vsak t velja $\widehat{q}_t(n, S) = \widehat{q}(n, [Y_t \in S]) = 1$, kjer so Y_t koordinatne projekcije. Zaradi desne zveznosti pa velja tudi:

$$\widehat{q}(n, [Y_t \in S \text{ za vsak } t]) = 1$$

To pa pomeni, da lahko markovsko jedro $\widehat{q}$ zožimo na prostor $S = \mathbb{N}_0$. Očitno je q Feller–Dynkinov proces s pripadajočimi lokalnimi jedri q_t in to je potem Poissonov proces. Kar smo definirali na začetku razdelka, pa je *Poissonov proces z začetkom v ničli.* To je torej slučajni proces X, ki ima za jedro Poissonov proces, začetna porazdelitev pa je skoncentrirana v 0. V tem primeru slučajna spremenljivka X_t predstavlja število dogodkov, ki so se zgodili v času od 0 do t in ima Poissonovo porazdelitev $\mathrm{Po}(\lambda t)$.

B.10 Brownovo gibanje

Brownovo gibanje je model za slučajno gibanje molekul v plinu, pa tudi za gibanje cen delnic in še mnoge druge kaotične pojave. Izkaže se, da ob privzemu dokaj smiselnih predpostavk nimamo kaj dosti izbire. Velja namreč naslednji izrek.

Izrek B.10.1. *Do linearne transformacije natančno obstaja natanko en Feller–Dynkinov proces q na $\mathbb{R}^n$, za katerega velja:*

(1) *Če je X slučajni proces z jedrom q, so njegove trajektorije skoraj gotovo zvezne (t. j. za skoraj vsak ω je preslikava $t \mapsto X_t(\omega)$ zvezna).*

(2) *Če je X slučajni proces z jedrom q, je to tudi $X + x$ za vsak $x \in \mathbb{R}^n$.*

(3) *Za kak (vsak) $t > 0$ je $\mathbf{E}(X_t) = 0$.*

Dokaz izreka bomo skicirali kasneje. Zdaj pa bomo konstruirali Feller–Dynkinov proces, za katerega se bo kasneje izkazalo, da ima zgornje lastnosti. Zahteva (2) pove, da mora veljati:

$$q(x, A) = q(0, A - x) \tag{B.10.1}$$

Iz nje po posledici B.9.2 sledi, da so za vsak $t \geq 0$ in vsak $n \in \mathbb{N}$ slučajne spremenljivke $X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}$, $k = 1, \ldots n$, neodvisne in enako porazdeljene. Če je $X_0 = 0$, je X_t njihova vsota. Smiselno bo privzeti, da velja centralni limitni izrek, iz katerega potem sledi, da je slučajna spremenljivka X_t porazdeljena normalno. Zaradi zahteve (3) mora biti $X_t \sim \mathrm{N}(0, \Sigma)$.

Oglejmo si variance slučajnih spremenljivk X_t! Če še kar naprej privzamemo, da je $X_0 = 0$, in upoštevamo formulo za vsoto varianc neodvisnih slučajnih spremenljivk, dobimo, da je $\mathrm{var}(X_t) = t\,\mathrm{var}(X_1)$ za vsak $t \in \mathbb{Q} \cap [0, \infty)$, zaradi desne zveznosti pa tudi za vsak $t \in [0, \infty)$. Brez škode za splošnost lahko privzamemo, da je $\mathrm{var}(X_1) = I$, kjer je I identiteta na $\mathbb{R}^n$. Vse ostale primere lahko dobimo iz tega z linearno transformacijo. Tako smo dobili, da je $X_t \sim \mathrm{N}(0, tI)$.

Ker smo privzeli $X_0 = 0$, bo to merilo za $q(0, A)$. Smiselno bo torej postaviti:

$$q_t(0, A) := \mathrm{N}(0, tI)(A) \tag{B.10.2}$$

od koder zaradi izotropnosti procesa sledi:

$$q_t(x, A) = \mathrm{N}(x, tI)(A) \tag{B.10.3}$$

Podobno kot pri Poissonovem procesu tudi tu zlahka preverimo, da za tako definirana jedra q_t velja enačba Chapmana in Kolmogorova. Tako bi že lahko konstruirali markovsko jedro. Če pa želimo konstruirati Feller–Dynkinov proces, si moramo spet ogledati pripadajoče linearne operatorje. Označimo $\gamma_n := \mathrm{N}_{\mathbb{R}_n}(0, I)$. Tedaj po posledici B.3.3 velja:

$$(P_t f)(x) = \int f\big(x + \sqrt{t}\, y\big)\, \gamma_n(dy)$$

Seveda velja $P_{t+s} = P_t P_s$. Če bi to hoteli neposredno preveriti, bi imeli kar nekaj vaje iz integracije. Namesto tega pa smo se sklicali na dejstvo, da je vsota neodvisnih normalno porazdeljenih slučajnih spremenljivk spet normalno porazdeljena.

Bralcu prepuščamo, naj preveri, da operatorji P_t slikajo $\mathcal{C}_0(\mathbb{R}^n)$ spet v $\mathcal{C}_0(\mathbb{R}^n)$. Po posledici B.8.6 obstaja natanko en Feller–Dynkinov proces $\widehat{q}$ na $\widehat{S} := \mathbb{R}^n \cup \{\infty\}$, ki mu pripada polgrupa operatorjev $\widehat{P}_t$, ki so ustrezne razširitve operatorjev P_t. Toda operatorji

P_t so že dobljeni iz jeder na prostoru $S = \mathbb{R}^n$ in ne na $\widehat{S}$. To pa pomeni, da slučajni proces z jedrom $\widehat{q}$, ki se začne v $\mathbb{R}^n$, skoraj gotovo ne uide v neskončnost. Potem pa lahko jedro $\widehat{q}$ zožimo na $\mathbb{R}^n$ in dobimo Feller–Dynkinov proces na $\mathbb{R}^n$. Temu potem pravimo Brownovo gibanje.

Pravkar konstruirani proces očitno izpolnjuje zahtevi (2) in (3). Zahtevo (1) bomo dokazali kasneje. Je pa zveznost trajektorij bistvena za enoličnost. Kasneje bomo namreč konstruirali proces rojevanja in umiranja, ki izpolnjuje zahtevi (2) in (3), ne izpolnjuje pa zahteve (1). Ta proces bo slučajni sprehod na $\mathbb{Z}$ v zveznem času in bo dobljen kot vsota dveh neodvisnih nasprotno predznačenih Poissonovih procesov.

Zaradi izotropnosti je navadno dovolj gledati Brownovo gibanje, ki se začne v točki 0. Dovolj je torej gledati porazdelitev $\mathbf{W} := \mathbf{P}_0$, torej porazdelitev $\mathbf{W}(A) := q(0, A)$. Tej porazdelitvi pravimo *Wienerjeva mera*. Iz nje potem rekonstruiramo jedro q po predpisu $q(x, A) = \mathbf{W}(A + x)$.

B.11 Generatorji operatorskih polgrup

Pri konstrukciji markovskih procesov smo že kar napredovali. Najprej je bilo treba poznati markovsko jedro iz $(S, \mathcal{S})$ v brezobličen prostor $(\overline{S}, \overline{\mathcal{S}})$, nato le še operatorsko polgrupo na še precej velikem prostoru $\flat(S, \mathcal{S})$, nazadnje pa le še na precej lepšem prostoru $\mathcal{C}(S)$ oz. $\mathcal{C}_0(S)$. Vendar pa je navadno celo operatorsko polgrupo težko podati in preveriti konsistentnost. Vendar pa se dajo tudi operatorske polgrupe konstruirati iz objektov, ki jih je mnogo lažje podati. Ti objekti so diferencialni ali podobni operatorji. Konstrukcijo operatorskih polgrup iz teh operatorjev nam da Hille–Yosidov izrek. Teorijo operatorskih polgrup povzemamo iz [10].

Definicija. *Operatorska polgrupa* je družina omejenih linearnih operatorjev P_t na Banachovem prostoru L, kjer je $t \in [0, \infty)$ in velja:

(1) $P_0 = I$.

(2) $P_{t+s} = P_t P_s$.

Polgrupa je *krepko zvezna*, če za vsak $f \in L$ velja $\lim_{t\downarrow 0} P_t f = f$. Polgrupa je *kontrakcijska*, če so njeni operatorji kontrakcije, t. j. če je $\|P_t\| \le 1$.

Do konca razdelka bo oznaka L pomenila Banachov prostor.

Opomba. Vsaka Feller–Dynkinova polgrupa je kontrakcijska krepko zvezna operatorska polgrupa.

Trditev B.11.1. *Naj bo $(P_t)_{t\in[0,\infty)}$ krepko zvezna operatorska polgrupa. Tedaj obstajata tak $M > 0$ in tak $\omega \in \mathbb{R}$, da za vsak $t \in [0, \infty)$ velja:*

$$\|P_t\| \le M e^{\omega t}$$

Opomba. Pri kontrakcijskih operatorskih polgrupah zgornja trditev seveda avtomatično velja. Vendar pa nam trditev pove tudi, da med krepko zveznimi operatorskimi polgrupami kontrakcijske niso zelo poseben primer. Brž ko je $(P_t)_{t\in[0,\infty)}$ krepko zvezna operatorska polgrupa in velja $\|P_t\| \le Me^{\omega t}$, operatorji $P'_t := e^{-\omega t}P_t$ tvorijo krepko zvezno operatorsko polgrupo, za katero velja ocena $\|P'_t\| \le M$. Če normo na prostoru L zamenjamo z ekvivalentno normo, definirano po predpisu:

$$\|f\|' := \sup_{t\in[0,\infty)} \|P'_t f\|$$

pa postane operatorska polgrupa $(P'_t)_{t\in[0,\infty)}$ kontrakcijska.

Dokaz trditve B.11.1. Najprej bomo pokazali, da obstajata tak $t_0 > 0$ in tak $M > 0$, da za vsak $t \in [0, t_0]$ velja $\|P_t\| \le M$. Če to ne bi bilo res, bi obstajalo tako zaporedje t_n, ki pada proti 0, da bi veljalo $\lim_{n\to\infty}\|P_{t_n}\| = \infty$. Toda zaradi krepke zveznosti je za vsak $f \in L$ zaporedje $P_{t_n} f$ omejeno in pridemo v protislovje s principom enakomerne omejenosti. Definirajmo $\omega := \frac{1}{t_0}\log M$. Naj bo $t \ge 0$. Obstajata tak $k \in \mathbb{N}_0$ in tak $s \in [0, t_0]$, da je $t = kt_0 + s$. Sledi:

$$\|P_t\| \le \|P_t^k P_s\| \le M\, M^{t/t_0} = Me^{\omega t}$$

■

Trditev B.11.2. *Naj bo $(P_t)_{t\in[0,\infty)}$ krepko zvezna operatorska polgrupa. Tedaj je za vsak $f \in L$ funkcija $t \mapsto P_t f$ zvezna funkcija iz $[0,\infty)$ v L.*

Dokaz. Naj bo $t \ge 0$ in naj bosta M in ω kot v trditvi B.11.1. Za $h \ge 0$ velja:

$$\|P_{t+h}f - P_t f\| = \|P_t(P_h f - f)\| \le Me^{\omega t}\|P_h f - f\|$$

za $0 \le h \le t$ pa velja:

$$\|P_{t-h}f - P_t f\| = \|P_{t-h}(P_h f - f)\| \le Me^{\omega t}\|P_h f - f\|$$

Od tod pa že sledi zveznost. ■

Definicija. *Linearni operator v Banachovem prostoru* L je (lahko tudi neomejen) linearni operator A iz $\mathcal{D}(A)$ v L, kjer je *definicijsko območje* $\mathcal{D}(A)$ linearni podprostor v L.

Operator A je *gosto definiran*, če je podprostor $\mathcal{D}(A)$ gost v L.

Graf linearnega operatorja A: $\mathcal{D}(A) \to L'$, kjer je $\mathcal{D}(A) \subset L$, je množica $\Gamma := \{(f, Af) \mid f \in \mathcal{D}(A)\}$. Operator A je *zaprt*, če je njegov graf zaprt v $L \times L'$ s produktno topologijo.

Zaloga vrednosti linearnega operatorja A je množica $\mathcal{R}(A) := \{Af \mid f \in \mathcal{D}(A)\}$.

Omejen linearni operator na L je omejen linearni operator iz L v L.

Opomba. Vsak omejen in povsod definiran linearni operator je zaprt. Drugače pa je linearni operator zaprt natanko tedaj, ko za vsako zaporedje $(f_n)_{n\in\mathbb{N}}$ elementov iz $\mathcal{D}(A)$, ki konvergira k f, velja: če zaporedje $(Af_n)_{n\in\mathbb{N}}$ konvergira h g, je $f \in \mathcal{D}(A)$ in $Af = g$.

Definicija. *(Infinitezimalni) generator* operatorske polgrupe $(P_t)_{t\in[0,\infty)}$ je (navadno neomejen) linearni operator G, definiran po predpisu:

$$Gf := \lim_{h\downarrow 0} \frac{P_h f - f}{h}$$

Definicijsko območje $\mathcal{D}(G)$ operatorja G so vsi elementi $f \in L$, za katere zgornja limita obstaja.

Trditev B.11.3. *Naj bo $(P_t)_{t\in[0,\infty)}$ krepko zvezna operatorska polgrupa z generatorjem G. Če je $f \in \mathcal{D}(G)$, je za vsak t tudi $P_t f \in \mathcal{D}(G)$ in velja:*

$$\frac{d}{dt} P_t f = G P_t f = P_t G f$$

Pri tem gre pri $t = 0$ le za desni, pri $t > 0$ pa za obojestranski odvod.

Dokaz. Za vsak $h > 0$ definirajmo $G_h := \frac{1}{h}(P_h - I)$. Velja:

$$G_h P_t f = \frac{1}{h}(P_{t+h} - P_t) f = P_t G_h f$$

Ker je $f \in \mathcal{D}(G)$ in je operator P_t omejen, desna stran konvergira proti $P_t G f$, ko gre h proti 0. Od tod pa že dobimo to, kar smo iskali, le da smemo namesto obojestranskega pisati le desni odvod. Da dokažemo, da smemo pisati tudi levi odvod, zgornjo enačbo zapišemo malo drugače. Če je $t > 0$, za $0 < h \le t$ velja:

$$\frac{1}{-h}(P_{t-h} - P_t) f = P_{t-h} G_h f$$

Dokazati moramo, da gre desna stran proti $P_t G f$. To pa sledi iz ocene:

$$\begin{aligned} \|P_{t-h} G_h f - P_t G f\| &\le \|P_{t-h}(G_h f - G f)\| + \|P_{t-h}(I - P_h) G f\| \le \\ &\le M e^{\omega t}\big(\|G_h f - G f\| + \|G f - P_h G f\|\big) \xrightarrow[h\downarrow 0]{} 0 \end{aligned}$$

kjer sta M in ω kot v trditvi B.11.1. ∎

Iz Feller–Dynkinove polgrupe $(P_t)_{t\in[0,\infty)}$ na prostoru S smo konstruirali normalizirano Feller–Dynkinovo polgrupo $(\widehat{P}_t)_{t\in[0,\infty)}$ na običajni kompaktifikaciji $\widehat{S}$ prostora S. Nove operatorje $\widehat{P}_t$ smo dobili tako, da smo predpisali $\widehat{P}_t 1 := 1$, kjer je konstanta 1 funkcija na $\widehat{S}$. Naj bo $\widehat{G}$ generator polgrupe $(\widehat{P}_t)_{t\in[0,\infty)}$. Očitno je $1 \in \mathcal{D}(\widehat{G})$ in velja $\widehat{G}1 = 0$. *Generator $\widehat{G}$ torej dobimo tako, da definicijskemu območju generatorja G prištejemo še prostor vseh konstant na $\widehat{S}$ in predpišemo $\widehat{G}1 = 0$.*

Definicija. Naj bo $\Delta = [a, b]$ omejen interval na realni osi. Dana naj bo preslikava $u\colon \Delta \to L$, kjer je L Banachov prostor. Izberimo delitev intervala $[a, b]$ s pripadajočimi vmesnimi točkami, naj bo torej $D = (t_0, x_1, t_1, x_2, \ldots x_n, t_n)$, kjer je $a = t_0 \le x_1 \le \ldots x_n \le t_n = b$, in ji priredimo *Riemannovo vsoto*:

$$R_D(u) := \sum_{k=1}^{n} u(x_k)(t_k - t_{k-1})$$

Za dano delitev D označimo $\operatorname{diam} D := \max_k (t_k - t_{k-1})$. Število R je *Riemannov integral* funkcije f na intervalu Δ, če za vsak $\varepsilon > 0$ obstaja tak $\delta > 0$, da za poljubno delitev D, za katero je $\operatorname{diam} D < \delta$, velja $\|R_D(u) - R\| < \varepsilon$.

Očitno je Riemannov integral funkcije u na Δ kvečjemu eden. Če obstaja, pravimo, da je preslikava u na Δ *integrabilna po Riemannu.* Njen integral označimo z $\int_a^b u(t)\,dt$ ali $\int_\Delta u(t)\,dt$.

Opomba. *Riemannov integral preslikave u obstaja natanko tedaj, ko za vsak $\varepsilon > 0$ obstaja tak $\delta > 0$, da za poljubni delitvi D_1 in D_2, za kateri je* $\operatorname{diam} D_1 < \delta$ *in* $\operatorname{diam} D_2 < \delta$, *velja, da je $\|R_{D_2}(u) - R_{D_1}(u)\| < \varepsilon$.* V tem primeru namreč izberemo poljubno zaporedje delitev D_n, katerih diametri konvergirajo proti 0. Zaporedje Riemannovih vsot $R_{D_n}(u)$ je potem Cauchyjevo, torej konvergentno. Ni težko preveriti, da je njegova limita ravno Riemannov integral preslikave u na Δ.

Definicija Riemannovega integrala preslikave v Banachov prostor je torej le prirejena definicija Riemannovega integrala realne funkcije. Tudi osnovne lastnosti so podobne, le da jih je nekoliko težje dokazati. Brez težav dokažemo naslednji trditvi.

Trditev B.11.4. *Če je $a < b$ in je funkcija u: $[a, b] \to L$, kjer je L Banachov prostor, po Riemannu integrabilna na $[a, b]$, je u omejena, funkcija $t \mapsto \|u(t)\|$ je tudi integrabilna in velja ocena:*

$$\left\| \int_a^b u(t)\,dt \right\| \leq \int_a^b \|u(t)\|\,dt \tag{B.11.1}$$

Trditev B.11.5. *Naj bo $a \leq b \leq c$. Dana naj bo preslikava u iz $[a, c]$ v Banachov prostor. Tedaj je u po Riemannu integrabilna na $[a, c]$ natanko tedaj, ko je po Riemannu integrabilna na $[a, b]$ in $[b, c]$. V primeru integrabilnosti veljajo formule:*

$$\int_a^c u(t)\,dt = \int_a^b u(t)\,dt + \int_b^c u(t)\,dt$$

$$\int_a^c u(t)\,dt = \lim_{b\uparrow c} \int_a^b u(t)\,dt$$

$$\int_a^c u(t)\,dt = \lim_{b\downarrow a} \int_b^c u(t)\,dt$$

Tako lahko zdaj integral $\int_a^b$ definiramo tudi za primer, ko je $a > b$: v tem primeru pač postavimo $\int_a^b := -\int_b^a$ in trditev še vedno velja.

Naj bo zdaj Δ poljuben interval na realni osi (odprt, zaprt, ali polodprt, omejen ali neomejen) in naj bo dana preslikava u: $\Delta \to L$. *Posplošeni Riemannov integral* funkcije f na Δ obstaja in je enak R, če je f po Riemannu integrabilna na poljubnem kompaktnem podintervalu intervala Δ in če za poljubno naraščajoče zaporedje kompaktnih intervalov Δ_n z unijo Δ zaporedje $\int_{\Delta_n} u(t)\,dt$ konvergira k R. Posplošeni Riemannov integral prav tako označimo z $\int_\Delta u(t)\,dt$ ali $\int_a^b u(t)\,dt$, kjer sta a in b krajišči intervala Δ.

Opomba. Iz trditve B.11.5 sledi, da je posplošeni Riemannov integral v resnici posplošitev običajnega: če obstaja običajni Riemannov integral, obstaja tudi posplošeni in integrala sta enaka. Poleg tega trditev B.11.5 velja tudi za posplošeni Riemannov integral, brž ko je u posplošeno integrabilna na (a, c). *Od zdaj naprej bodo za nas integrabilne preslikave tiste, ki so posplošeno integrabilne po Riemannu.*

Naj bo Δ interval na realni osi. S $\mathcal{C}_L(\Delta)$ označimo prostor vseh zveznih preslikav iz Δ v L, s $\mathcal{C}^1_L(\Delta)$ pa prostor vseh preslikav iz Δ v L, ki so zvezne na Δ in zvezno odvedljive v notranjosti intervala Δ.

Trditev B.11.6. *Posplošeni Riemannov integral ima naslednje lastnosti.*

(1) *Če je $u \in \mathcal{C}_L(\Delta)$ in je $\int_\Delta \|u(t)\| < \infty$, je u integrabilna na Δ in spet velja ocena:*

$$\left\| \int_a^b u(t)\, dt \right\| \leq \int_a^b \|u(t)\|\, dt$$

(2) *Naj bo $A\colon \mathcal{D}(A) \to L'$, kjer je $\mathcal{D}(A) \subset L$, zaprt linearni operator. Naj bo Δ interval na $\mathbb{R}$ in $u\colon \Delta \to L$ integrabilna preslikava. Za vsak $t \in \Delta$ naj bo $u(t) \in \mathcal{D}(A)$ in naj bo tudi preslikava Au integrabilna. Tedaj je $\int_\Delta u(t)\, dt \in \mathcal{D}(A)$ in velja formula:*

$$A \int_\Delta u(t)\, dt = \int_\Delta Au(t)\, dt$$

(3) *Če je $u \in \mathcal{C}^1_L([a, b])$, velja:*

$$\int_a^b \frac{d}{dt} u(t)\, dt = u(b) - u(a)$$

(4) *Naj bo $a < b$ in u zvezna in integrabilna na $[a, b)$. Potem za vsak $t \in [a, b)$ velja:*

$$\frac{d}{dt} \int_a^t u(s)\, ds = u(t)$$

Pri tem je za $t = a$ mišljen desni, za $a < t < b$ pa obojestranski odvod.

Skica dokaza. Točko (1) najprej dokažemo za primer, ko je interval Δ kompakten. V tem promeru imamo opravka z običajnim Riemannovim integralom. Dokazati moramo, da se Riemannovi vsoti po delitvah, ki imata majhna diametra, ne razlikujeta preveč. To storimo tako, da najprej poiščemo delitev, ki je finejša od obeh (za vmesne točke se ne menimo) in upoštevamo enakomerno zveznost. Ta korak je malo bolj zoprn kot pri integralu realne funkcije, ker si ne moremo pomagati z Darbouxovimi vsotami. Pri dokazovanju posplošene integrabilnosti po Riemannu pa je glavno orodje ocena (B.11.1).

V točki (2) upoštevamo, da je vsak (tudi posplošen) Riemannov integral limita primernih Riemannovih vsot. Riemannovo vsoto smemo zamenjati z operatorjem A, saj je le-ta linearen. Torej velja $A\, R_D(u) = R_D(Au)$. Kratek premislek pokaže, da lahko izberemo tako zaporedje D_n delitev primernih podintervalov intervala Δ, da gre $R_{D_n}(u)$ proti $\int_\Delta u(t)\, dt$ in $R_{D_n}(Au) = AR_{D_n}(u)$ proti $\int_\Delta Au(t)\, dt$. Ko naredimo limito in upoštevamo, da je A zaprt, dobimo, kar smo iskali.

Točko (3) pa tudi najprej dokažemo za primer, ko je u odvedljiva na vsem intervalu $[a, b]$ (t. j. v a obstaja desni, v b pa levi odvod). Dokaz je spet malo bolj zoprn kot v realnem primeru, ker ne moremo uporabiti Lagrangeovega izreka. Naj bo $\varepsilon > 0$ in $\delta > 0$. Za vsak $x \in [a, b]$ obstaja tak $\delta_x > 0$, da za vsak t, za katerega je $|t - x| < \delta_x$, velja

$\|\frac{u(t)-u(x)}{t-x} - u'(x)\| < \varepsilon$. Od tod sledi, da za poljubna t_1 in t_2, za katera je $x - \delta_x < t_1 \leq x \leq t_2 < x + \delta_x$, velja ocena:

$$\|u(t_2) - u(t_1) - u'(x)(t_2 - t_1)\| \leq \varepsilon(t_2 - t_1)$$

Brez škode za splošnost lahko postavimo $\delta_x < \delta$. Pokritje intervala $[a, b]$ z intervali $(x - \delta_x, x + \delta_x)$ ima končno podpokritje, iz katerega dobimo delitev D intervala $[a, b]$ z diametrom, manjšim od δ, za katero velja ocena $\|R_D(u') - (u(b) - u(a))\| \leq \varepsilon(b - a)$. Iz točke (1) potem sledi to, kar smo hoteli dokazati. Splošni primer pa dobimo z limitiranjem.

Pri točki (4) pa je dovolj preveriti, da velja:

$$\lim_{h \to 0} \frac{1}{h} \int_t^{t+h} u(s)\, ds = u(t)$$

Če je $a < t + h < b$, lahko zgornji integral obravnavamo kot navaden Riemannov integral. Za vsak $\varepsilon > 0$ obstaja tak $\delta > 0$, da je za vsak s, za katerega je $|s - t| < \delta$, tudi $\|u(s) - u(t)\| < \varepsilon$. Če je $0 < h < \delta$ in je D delitev intervala $[t, t+h]$, je potem $\left\|\frac{1}{h} R_D(u) - u(t)\right\| < \varepsilon$, potem pa je tudi $\left\|\frac{1}{h} \int_t^{t+h} u(t)\, dt - u(t)\right\| < \varepsilon$. Slednjo oceno pa lahko na enak način dokažemo tudi za $-\delta < h < 0$. Iz te ocene pa že sledi to, kar smo hoteli dokazati. ■

Opomba. Če je L prostor realnih funkcij s supremum normo, so evaluacijski funkcionali $x \mapsto f(x)$ gotovo omejeni. Iz prejšnje trditve potem sledi:

$$\left(\int_\Delta u(t)\, dt\right)(x) = \int_\Delta u(t)(x)\, dt \tag{B.11.2}$$

Integral se torej prevede na običajni Riemannov integral realne funkcije.

Trditev B.11.7 (Dynkinova formula). *Naj bo $(P_t)_{t \in [0,\infty)}$ krepko zvezna operatorska polgrupa z generatorjem G. Naj bo $f \in L$ in $t \geq 0$. Tedaj je $\int_0^t P_s f\, ds \in \mathcal{D}(G)$ in velja:*

$$P_t f - f = G \int_0^t P_s f\, ds$$

Če pa je $f \in \mathcal{D}(G)$, velja še:

$$P_t f - f = \int_0^t G P_s f\, ds = \int_0^t P_s G f\, ds$$

Dokaz. Iz točke (2) prejšnje trditve in trditve B.11.5 sledi, da za vsak $h > 0$ velja:

$$\begin{aligned}
\frac{1}{h}(P_h - I) \int_0^t P_s f\, ds &= \frac{1}{h} \int_0^t (P_{s+h} f - P_s f)\, ds = \\
&= \frac{1}{h} \left(\int_h^{t+h} P_s f\, ds - \int_0^t P_s f\, ds\right) = \\
&= \frac{1}{h} \int_t^{t+h} P_s f\, ds - \frac{1}{h} \int_0^h P_s f\, ds
\end{aligned}$$

Ko gre h proti 0, iz definicije generatorja in točke (4) prejšnje trditve sledi prva enačba, ki jo je bilo treba dokazati. Druga enačba pa takoj sledi iz trditve B.11.3 in točke (3) prejšnje trditve. ∎

Opomba. Na prvi pogled bi se zdelo, da lahko drugi del trditve izpeljemo kar iz točke (2) prejšnje trditve. Vendar pa tega ne moremo storiti, ker še ne vemo, da je generator G zaprt. To pa je naslednja stvar, ki jo bomo izpeljali.

Posledica B.11.8. *Generator krepko zvezne operatorske polgrupe je gosto definiran in zaprt.*

Dokaz. Po točki (1) prejšnje trditve je $f_t := \int_0^t P_s f \, ds \in \mathcal{D}(G)$ za vsak $t > 0$ in vsak $f \in L$. Ker je $\lim_{t \downarrow 0} \frac{f_t}{t} = f$, je G res gosto definiran. Naj zdaj zaporedje f_n konvergira proti f, zaporedje Gf_n pa proti g. Po točki (2) prejšnje trditve je $P_t f_n - f_n = \int_0^t P_s G f_n \, ds$. Pošljimo najprej n proti neskončno! Leva stran očitno konvergira proti $P_t f - f$. Iz enakomerne omejenosti operatorjev P_s (trditev B.11.1) in ocene (B.11.1) pa dobimo, da gre desna stran proti $\int_0^t P_s g \, ds$. Zdaj pa enačbo $P_t f - f = \int_0^t P_s g \, ds$ delimo s t, pošljemo t proti 0 in ob upoštevanju točke (4) trditve B.11.6 dobimo, da je $Gf = g$, od tod pa sledi, da je generator g res zaprt. ∎

Definicija. Naj bo A zaprt linearni operator v Banachovem prostoru L. Operator A je *obrnljiv*, če je A bijektivna preslikava iz $\mathcal{D}(A)$ na L. *Resolventna množica* $\rho(A)$ operatorja A je množica tistih realnih števil λ, za katera je operator $\lambda I - A$ obrnljiv.

Opomba. Iz izreka o zaprtem grafu sledi, da je inverz obrnljivega operatorja omejen linearni operator na L.

Trditev B.11.9. *Naj bo* $(P_t)_{t \in [0,\infty)}$ *krepko zvezna kontrakcijska operatorska polgrupa z generatorjem* G. *Tedaj je* $(0, \infty) \subset \rho(G)$ *in za vsak* $f \in L$ *in vsak* $\lambda > 0$ *velja:*

$$(\lambda I - G)^{-1} f = \int_0^\infty e^{-\lambda t} P_t f \, dt$$

Opomba. Integral na desni ni nič drugega kot Laplaceova transformiranka preslikave $t \mapsto P_t f$.

Dokaz bomo opustili, bralec ga lahko najde v [10] ali pa sam dokaže za vajo. Pri Chen–Steinovi metodi pa nas bo bolj zanimal primer, ko je $\lambda = 0$. Tu trditev, tako kot je napisana, ne velja, potrebno je postaviti dodatne pogoje tako pri operatorski polgrupi kot tudi pri elementu f.

Opomba. Trditev je mogoče formulirati tudi za splošne krepko zvezne operatorske polgrupe (glej opombo k trditvi B.11.1).

Definicija. Linearni operator A v Banachovem prostoru L je *disipativen*, če za vsak $f \in \mathcal{D}(A)$ in vsak $\lambda > 0$ velja:

$$\|\lambda f - Af\| \geq \lambda \|f\|$$

Naj bo $\lambda > 0$. Iz ocene:

$$\left\| \int_0^\infty e^{-\lambda t} P_t f \, dt \right\| \leq \int_0^\infty e^{-\lambda t} \|P_t f\| \, dt \leq \frac{\|f\|}{\lambda}$$

in prejšnje trditve takoj sledi, da je generator krepko zvezne kontrakcijske operatorske polgrupe disipativen.

Navedli smo že kar nekaj lastnosti generatorjev krepko zveznih kontrakcijskih operatorskih polgrup. Nastane vprašanje, ali te lastnosti že karakterizirajo naše generatorje. Odgovor na to nam da Hille–Yosidov izrek.

Izrek B.11.10 (Hille, Yosida). *Linearni operator G v Banachovem prostoru je generator krepko zvezne kontrakcijske operatorske polgrupe natanko tedaj, ko je gosto definiran, disipativen in ko je za kak $\lambda > 0$ operator $\lambda I - G$ surjektiven. Vsak linearni operator generira največ eno operatorsko polgrupo.*

Dokaz tega izreka bomo opustili, bralec ga spet najde v [10] (izrek 2.6 in trditev 2.9). Bolj nas bodo zanimale njegove daljnosežne posledice. Hille–Yosidov izrek nam skupaj s posledico B.8.6 pove, da nam vsak linearni operator v prostoru $\mathcal{C}_0(S)$, ki izpolnjuje pogoje izreka, določa Feller–Dynkinov proces na kompaktifikaciji $\widehat{S}$. V naslednjem razdelku bomo navedli nekaj zgledov.

B.12 Zgledi generatorjev

Zdaj, ko smo videli, da obstaja bijektivna korespondenca med Feller–Dynkinovimi procesi in generatorji, bi kazalo navesti nekaj zgledov. Izračunali bomo generatorja Poissonovega procesa in Brownovega gibanja, konstruirali pa bomo tudi dva nova razreda Feller–Dynkinovih procesov.

1. Poissonov proces. Spomnimo se, da so elementi operatorske polgrupe, ki pripada Poissonovemu procesu, podani po predpisu:

$$(P_t f)(n) := \sum_{k=0}^{\infty} \frac{(t\lambda)^k e^{-t\lambda}}{k!} f(n+k)$$

Sledi:

$$\begin{aligned} (P_t f)(n) - f(n) &= \sum_{k=1}^{\infty} \frac{(t\lambda)^k e^{-t\lambda}}{k!} \big(f(n+k) - f(n)\big) \\ \frac{(P_t f)(n) - f(n)}{t} &= \lambda e^{-t\lambda}\big(f(n+1) - f(n)\big) + \sum_{k=2}^{\infty} \frac{t^{k-1}\lambda^k e^{-t\lambda}}{k!} \big(f(n+k) - f(n)\big) \end{aligned}$$

Ko gre t proti 0, gre prvi člen na desni proti $\lambda\big(f(n+1) - f(n)\big)$, vrsta na desni pa proti 0, saj za $0 < t \le 1$ velja ocena:

$$\left| \sum_{k=2}^{\infty} \frac{t^{k-1}\lambda^k e^{-t\lambda}}{k!} \big(f(n+k) - f(n)\big) \right| \le 2te^{\lambda} \|f\|$$

kjer je $\|\cdot\|$ kot ponavadi supremum norma. Generator Poissonovega procesa je torej podan po predpisu:

$$(Gf)(n) = \lambda\big(f(n+1) - f(n)\big)$$

Opazimo, da je generator povsod definiran in omejen. To pa v splošnem še zdaleč ne bo res.

Oglejmo si še malo pomen našega generatorja! Če je X Poissonov proces in skoraj gotovo $X_0 = n$, je $(P_t f)(n) = \mathbf{E}(f(X_t))$. Torej je $(Gf)(n)$ desni odvod funkcije $t \mapsto \mathbf{E}(f(X_t))$ v točki 0. Pri $t = 0$ je seveda $\mathbf{E}(X_t)) = f(n)$, pri malo večjem t pa približno velja $\mathbf{E}(X_t) = (1 - \lambda t)f(n) + \lambda t f(n+1)$, kar ne pomeni nič drugega kot to, da je $X_t - X_0$ približno Bernoullijeva slučajna spremenljivka. Približno torej velja, da X_t z verjetnostjo λt skoči v $n + 1$, sicer pa ostane v n. Parameter λ meri pogostnost skokov.

2. Uteženi slučajni sprehodi v zveznem času. Ti procesi so posplošitev Poissonovega procesa in eden od njih tiči v ozadju Poissonove aproksimacije po Chen–Steinovi metodi. Pri Poissonovem procesu iz vsake točke $n \in \mathbb{Z}$, kjer trenutno smo, s težnjo λ skočimo v $n + 1$. Pri uteženem slučajnem sprehodu pa s težnjo a_n skočimo v $n + 1$ in s težnjo b_n v $n - 1$. Za generator našega procesa bomo torej postavili:

$$(Gf)(n) := a_n\big(f(n+1) - f(n)\big) + b_n\big(f(n-1) - f(n)\big) \tag{B.12.1}$$

Če je $b_0 = 0$, lahko za množico stanj S tako kot pri Poissonovem procesu postavimo $\mathbb{N}_0$. Takim slučajnim procesom (oziroma njihovim jedrom) bomo rekli *procesi rojevanja in umiranja*, saj si lahko stanje predstavljamo kot velikost populacije.

V nasprotju s Poissonovim procesom tokrat generator ne bo nujno povsod definiran in tudi omejen ne bo. Za definicijsko območje operatorja G seveda postavimo vse funkcije $f \in \mathcal{C}_0(\mathbb{Z})$, za katere zgoraj definirana funkcija Gf tudi pripada $\mathcal{C}_0(\mathbb{Z})$. Očitno je G gosto definiran, saj so v definicijskem območju tudi vse funkcije s kompaktnim nosilcem.

Seveda je smiselno privzeti, da je $a_n \geq 0$ in $b_n \geq 0$. V tem primeru bo operator G disipativen. Vsaka funkcija $f \in \mathcal{C}_0(\mathbb{Z})$ doseže maksimum svoje absolutne vrednosti v neki točki $n \in \mathbb{Z}$. Naj bo $f(n) = M$. Naj bo najprej $M \geq 0$. Potem je očitno $(Gf)(n) \leq 0$, zato za vsak $\lambda > 0$ velja $\lambda f(n) - (Gf)(n) \geq \lambda M = \|f\|$. V primeru, ko je $M < 0$, pa f zamenjamo z $-f$ in dobimo isto oceno. Operator G je torej res disipativen.

Dosti težje vprašanje pa je, ali naš operator izpolnjuje še zadnji pogoj Hille–Yosidovega izreka, t. j., da je za kak (vsak) $\lambda > 0$ operator $\lambda I - G$ surjektiven. Privzemimo, da je to res. Tedaj operator G po posledici B.8.6 določa Feller–Dynkinov proces na $\mathbb{Z} \cup \{\infty\}$, ki mu pravimo *utežen slučajni sprehod v zveznem času.* V primeru, ko je $b_0 = 0$ in množico stanj omejimo na $\mathbb{N}_0$, pa temu Feller–Dynkinovemu procesu pravimo *proces rojevanja in umiranja.*

Zanimivo vprašanje je, kdaj lahko tak proces zožimo na $\mathbb{Z}$, torej, kdaj nam proces skoraj gotovo ne uide v neskončnost. Odgovor bi poznali, če bi poznali operatorsko polgrupo $(P_t)_{t\in[0,\infty)}$. Čeprav je dokaz Hille–Yosidovega izreka konstruktiven, pa je konstrukcija dovolj zapletena, da je iz nje zelo težko razbirati lastnosti operatorjev P_t. Vprašanje, kdaj proces uide v neskončnost, bomo pustili odprto.

Zgled 1. Poissonov proces si lahko predstavljamo kot proces čistega rojevanja: populacija se s konstantno intenziteto veča. Oglejmo si zdaj še proces čistega umiranja. Pogostnost umiranja pa tu ne bo konstantna, temveč bo sorazmerna velikosti populacije. Postavili bomo torej $a_n = 0$ in $b_n = n$ za $n \in \mathbb{N}_0$. Ker je $b_0 = 0$, lahko za množico stanj postavimo kar $\mathbb{N}_0$. Generator našega procesa bo torej definiran po predpisu:

$$(Gf)(n) := \begin{cases} n\big(f(n-1) - f(n)\big) & n > 0 \\ 0 & n = 0 \end{cases}$$

Pokažimo, da G izpolnjuje pogoje Hille–Yosidovega izreka. Pokazati je treba le, da obstaja tak $\lambda > 0$, da za vsak $f \in \mathcal{C}_0(\mathbb{N}_0)$ obstaja taka funkcija $g \in \mathcal{C}_0(\mathbb{N}_0)$, da velja:

$$\lambda g - Gg = f \tag{B.12.2}$$

oziroma:

$$\begin{aligned} \lambda g(0) &= f(0) \\ (\lambda + n)g(n) - ng(n-1) &= f(n) \end{aligned}$$

Če naredimo substitucijo:

$$h(n) := \frac{(\lambda+1)(\lambda+2)\ldots(\lambda+n)}{n!} g(n)$$

dobimo enostaven sistem:

$$\begin{aligned} \lambda h(0) &= f(0) \\ h(n) - h(n-1) &= f(n)\frac{(\lambda+1)\ldots(\lambda+n-1)}{n!} \end{aligned}$$

katerega rešitev je:

$$h(n) = \frac{1}{\lambda}\sum_{k=0}^{n} f(k)\frac{\lambda(\lambda+1)\ldots(\lambda+k-1)}{k!}$$

oziroma:

$$g(n) = \sum_{k=0}^{n} f(k)\frac{(k+1)(k+2)\ldots n}{(\lambda+k)(\lambda+k+1)\ldots(\lambda+n)}$$

Dokazati moramo, da je $g \in \mathcal{C}_0(\mathbb{N}_0)$, če je $f \in \mathcal{C}_0(\mathbb{N}_0)$. Z nekaj truda bi lahko neposredno preverili, da to velja za vsak $\lambda > 0$. Vendar pa nam tega ni treba preveriti za vsak $\lambda > 0$, dovolj je preveriti za en sam λ. Za $\lambda = 1$ dobi formula za funkcijo g precej preglednejšo obliko:

$$g(n) = \frac{1}{n+1}\sum_{k=0}^{n} f(k)$$

Limita tega izraza, ko gre n proti neskončno, je ravno Cesarova limita, ki je enaka limiti $\lim_{n\to\infty} f(n)$, brž ko slednja obstaja. Brž ko je torej $f \in \mathcal{C}_0(\mathbb{N}_0)$, je tudi $g \in \mathcal{C}_0(\mathbb{N}_0)$, torej je enačba (B.12.2) rešljiva, zato G izpolnjuje pogoje Hille–Yosidovega izreka in določa Feller–Dynkinov proces.

V prejšnjem zgledu smo "na roko" preverili, da G izpolnjuje pogoje Hille–Yosidovega izreka, in to tako, da smo kar neposredno rešili enačbo (B.12.2). Navadno pa se tega ne da narediti. Izpeljimo torej nekaj skromnih zadostnih pogojev za to, da operator G, definiran po formuli (B.12.1), izpolnjuje pogoje Hille–Yosidovega izreka. Ključ do njih je naslednja trditev.

Trditev B.12.1. *Naj bo L Banachov prostor in naj bo A: $\mathcal{D}(A) \to L$ obrnljiv linearni operator z omejenim inverzom. Naj bo B: $L \to L$ omejen linearni operator. Če je $\|B\| < \|A^{-1}\|^{-1}$, je tudi operator $A + B$ obrnljiv.*

Dokaz. Ker je $\|A^{-1}B\| \leq \|A^{-1}\| \, \|B\| < 1$ in je L Banachov prostor, je operator $I + A^{-1}B$ obrnljiv (njegov inverz dobimo z Neumannovo vrsto). Naj bo torej $C := (I + A^{-1}B)^{-1}$. Najprej na $\mathcal{D}(A)$ velja:

$$CA^{-1}(A + B) = C(I + A^{-1}B) = I$$

torej je CA^{-1} levi inverz operatorja $A + B$. Da dokažemo, da je tudi desni inverz, najprej opazimo, da je $\mathcal{D}(A)$ invariantni podprostor operatorja $I + A^{-1}B$, torej tudi invariantni podprostor operatorja C. Potem pa lahko zapišemo:

$$(A + B)CA^{-1} = A(I + A^{-1}B)CA^{-1} = AA^{-1} = I$$

torej je CA^{-1} tudi desni inverz. ∎

Trditev B.12.2. *Naj bodo koeficienti a_n in b_n omejeni. Tedaj operator G izpolnjuje pogoje Hille–Yosidovega izreka.*

Dokaz. Če so koeficienti a_n in b_n omejeni, je omejen tudi operator G. Po prejšnji trditvi je potem operator $\lambda I - G$ obrnljiv, brž ko je $\lambda > \|G\|$. Potem pa G že izpolnjuje pogoje Hille–Yosidovega izreka. ∎

Zgornja trditev pa se da še posplošiti. Včasih lahko generator razbijemo na dva operatorja, od katerih eden izpolnjuje pogoje Hille–Yosidovega izreka, drugi pa je omejen.

Trditev B.12.3. *Naj bo $a_n = a'_n + a''_n$ in $b_n = b'_n + b''_n$. Naj bo tudi $a'_n \geq 0$ in $b'_n \geq 0$ (že vseskozi pa privzemamo tudi $a_n \geq 0$ in $b_n \geq 0$). Naj operator G', definiran po predpisu:*

$$(G'f)(n) := a'_n\big(f(n+1) - f(n)\big) + b'_n\big(f(n-1) - f(n)\big)$$

izpolnjuje pogoje Hille–Yosidovega izreka in naj bodo koeficienti a''_n in b''_n omejeni. Tedaj tudi operator G, definiran po formuli (B.12.1), izpolnjuje pogoje Hille–Yosidovega izreka.

Dokaz. Iz omejenosti koeficientov a''_n in b''_n sledi, da je tudi operator $G'' := G - G'$ omejen. Iz disipativnosti operatorja $\lambda I - G'$ pa sledi, da je $\|(\lambda I - G)^{-1}\| \leq \frac{1}{\lambda}$. Brž ko je torej $\lambda > \|G''\|$, je $\|G''\| < \|(\lambda I - G)^{-1}\|^{-1}$, potem pa je po trditvi B.12.1 operator $\lambda I - G = \lambda I - G' - G''$ obrnljiv, torej tudi G izpolnjuje pogoje Hille–Yosidovega izreka. ∎

Zgled 2. Sestavimo procesa rojevanja in umiranja, ki smo ju spoznali, v en sam proces. Postavimo torej $a_n := \lambda$ in $b_n := n$, množico stanj pa seveda omejimo na $\mathbb{N}_0$. Pokažimo, da operator G, ki ga po formuli (B.12.1) določajo ti koeficienti, izpolnjuje pogoje Hille–Yosidovega izreka. Generator G lahko zapišemo kot vsoto $G' + G''$, kjer je G' generator procesa čistega umiranja, G'' pa generator procesa čistega rojevanja. Videli smo, da G' izpolnjuje pogoje Hille–Yosidovega izreka, G'' pa je omejen. Po prejšnji trditvi potem tudi G izpolnjuje pogoje Hille–Yosidovega izreka, torej določa Feller–Dynkinov proces. Ta proces ima ravnovesno porazdelitev, in sicer je to Poissonova porazdelitev (glej zgled 2 na strani 137). In v resnici je ta proces ozadje Poissonove aproksimacije po Chen–Steinovi metodi.

3. Brownovo gibanje. Operatorska polgrupa Brownovega gibanja je (do linearne transformacije natančno) podana po predpisu:

$$(P_t f)(x) = \int f\big(x + \sqrt{t}\, y\big)\, \gamma_n(dy) \tag{B.12.3}$$

Da je to krepko zvezna operatorska polgrupa, sledi iz tega, da je dobljena iz konkretnega Feller–Dynkinovega procesa. Ključen za krepko zveznost je izrek B.8.4. Seveda pa bi se dala krepka zveznost preveriti tudi neposredno, s tehniko epsilon–delta.

Naj bo $f \in \mathcal{C}_0(\mathbb{R}^n)$. Če želimo narediti limito, ko gre t proti 0, moramo f razviti v Taylorjevo vrsto. Naj bo f dvakrat zvezno odvedljiva in $t \ge 0$. Potem velja:

$$f\big(x + \sqrt{t}\, y\big) = f(x) + \sqrt{t} \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)\, y_i + \frac{t}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i\, \partial x_j}(\xi)\, y_i y_j$$

Tu je $\xi = x + \vartheta \sqrt{t}\, y$, kjer je $0 \le \vartheta \le 1$. Če želimo, da bo $f \in \mathcal{D}(G)$, bodo morali iti drugi parcialni odvodi v točki ξ proti drugim parcialnim odvodom v točki 0, in sicer enakomerno po vseh $x \in \mathbb{R}^n$. Naj bodo vsi drugi parcialni odvodi funkcije f v $\mathcal{C}_0(\mathbb{R}^n)$. V tem primeru so le-ti tudi enakomerno zvezni. Če torej definiramo:

$$\varphi(r) := \sup_{\substack{\|x-y\|<r \\ 1 \le i,j \le n}} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(x) - \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \right|$$

velja $\lim_{r \to 0} \varphi(r) = 0$. Sledi:

$$f\big(x + \sqrt{t}\, y\big) = f(x) + \sqrt{t} \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)\, y_i + \frac{t}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i\, \partial x_j}(x)\, y_i\, y_j + R(x, y, t)$$

kjer je:

$$|R(x, y, t)| \le \frac{t}{2} \varphi\big(\sqrt{t} \|y\|\big) \sum_{i=1}^n \sum_{j=1}^n |y_i y_j|$$

Očitno je $\int \gamma_n(dy) = 1$, $\int y_i\, \gamma_n(dy) = 0$ in $\int y_i^2 \gamma_n(dy) = 1$. Za $i \neq j$ pa je $\int y_i y_j\, \gamma_n(dy) = 0$. Sledi:

$$(P_t f)(x) = f(x) + \frac{t}{2} (\Delta f)(x) + \int R(x, y, t)\, \gamma_n(dy)$$

Z Δ smo označili Laplaceov operator. Sledi:

$$\left\| \frac{P_t f - f}{t} - \frac{1}{2} \Delta f \right\| \le \frac{1}{2} \int \varphi\big(\sqrt{t}\, y\big) \sum_{i=1}^n \sum_{j=1}^n |y_i y_j|\, \gamma_n(dy) \tag{B.12.4}$$

Ker so drugi parcialni odvodi v $\mathcal{C}_0(\mathbb{R}^n)$, so tudi omejeni. Torej je tudi funkcija φ navzgor omejena. Potem pa lahko pri limitiranju, ko gre t proti 0, uporabimo izrek o dominirani konvergenci, torej desna in zato tudi leva stran konvergira proti 0. Dokazali smo naslednji izrek.

Izrek B.12.4. *Naj bo $(P_t)_{t\in[0,\infty)}$ operatorska polgrupa, ki pripada Brownovemu gibanju, in naj bo:*

$$D := \left\{ f \in \mathcal{C}_0(\mathbb{R}^n) \cap \mathcal{C}^2(\mathbb{R}^n) \;\middle|\; (\forall i,j)\, \tfrac{\partial^2 f}{\partial x_i \partial x_j} \in \mathcal{C}_0(\mathbb{R}^n) \right\} \tag{B.12.5}$$

Naj bo G generator operatorske polgrupe $(P_t)_{t\in[0,\infty)}$. Tedaj je $D \subset \mathcal{D}(G)$ in za vsak $f \in D$ velja $Gf = \frac{1}{2}\Delta f$.

Generator G torej zaenkrat poznamo na D. Nastane seveda vprašanje, ali je G natančno določen s svojo zožitvijo na D. Natančneje, ker vemo, da je G zaprt operator, se je smiselno vprašati, ali je G zaprtje svoje zožitve na D, torej zaprtje operatorja $\frac{1}{2}\Delta$. Odgovor je pritrdilen, ključ do njega pa je naslednja trditev, ki je dokazana v [10] kot trditev 3.3.

Trditev B.12.5. *Naj bo G generator krepko zvezne kontrakcijske operatorske polgrupe $(P_t)_{t\in[0,\infty)}$ na Banachovem prostoru L. Naj bo D_0 gost podprostor v L in naj bo $D_0 \subset D \subset \mathcal{D}(G)$. Če za vsak $t \geq 0$ operator P_t slika D_0 v D, je G zaprtje svoje zožitve na D.* ■

V našem primeru bomo za prostor D_0 vzeli prostor $\mathcal{C}^2_\kappa(\mathbb{R}^n)$, t. j. prostor vseh dvakrat zvezno odvedljivih funkcij s kompaktnim nosilcem. Očitno je ta prostor gost v $\mathcal{C}_0(\mathbb{R}^n)$. Dokazati moramo, da operatorji P_t slikajo D_0 v D. Preden pa gremo to dokazat, dokažimo še dve lemi, ki ju ne bomo potrebovali le tu, zato ju bomo izpeljali za splošnejši primer.

Definicija. Funkcija $f\colon \mathbb{R}^n \to \mathbb{R}$ je *eksponentno omejena*, če obstajata tak $M > 0$ in tak $\alpha \in \mathbb{R}$, da je $|f(x)| \leq M e^{\alpha\|x\|}$ za vsak $x \in \mathbb{R}^n$.

Očitno so operatorji P_t s formulo (B.12.3) dobro definirani tudi na prostoru vseh eksponentno omejenih funkcij (funkcije lahko naraščajo še dosti hitreje, toda več ne bomo nikjer potrebovali).

Lema B.12.6. *Naj bodo P_t operatorji, definirani po formuli (B.12.3), delujejo pa naj na prostoru eksponentno omejenih funkcij na $\mathbb{R}^n$. Naj bo $r \in \mathbb{N}$, naj bo $f \in \mathcal{C}^r(\mathbb{R}^n)$ in naj bo funkcija f skupaj z vsemi svojimi parcialni odvodi do vključno reda r eksponentno omejena. Naj bo ∂ kak parcialni odvod reda r. Tedaj za vsak $t \in [0,\infty)$ element $P_t f$ pripada $\mathcal{C}^r(\mathbb{R}^n)$ in velja $\partial P_t f = P_t \partial f$.*

Opomba. V resnici je $P_t f$ neskončno gladka funkcija že za poljubno merljivo eksponentno omejeno funkcijo f, brž ko je $t > 0$.

Dokaz leme. Dovolj je dokazati za $r = 1$, saj lahko splošni primer izpeljemo z indukcijo. Naj bo torej $f \in \mathcal{C}^1(\mathbb{R}^n) \cap L^1(\gamma_n)$ in za vse $i = 1,\dots n$ naj bo $\frac{\partial f}{\partial x_i} \in L^1(\gamma_n)$. Označimo $e_i := (0,\dots 1,\dots 0)$, kjer je enica na i-tem mestu. Potem velja:

$$\begin{aligned}\left(\frac{\partial}{\partial x_i} P_t f\right)(x) &= \lim_{h\to 0} \frac{(P_t f)(x+h) - (P_t f)(x)}{h} = \\ &= \lim_{h\to 0} \int \frac{f\big(x + he_i + \sqrt{t}\,y\big) - f\big(x + \sqrt{t}\,y\big)}{h}\, \gamma_n(dy)\end{aligned}$$

Ključno vprašanje je, ali lahko uporabimo izrek o dominirani konvergenci. Toda parcialni odvod $\frac{\partial f}{\partial x_i}$ je eksponentno omejen, torej obstajata taka M in α, da je $\left|\frac{\partial f}{\partial x_i}(x)\right| \le M e^{\alpha\|x\|}$. Za $|h| \le 1$ lahko potem integrand ocenimo po Lagrangeovem izreku:

$$\left|\frac{f\big(x + he_i + \sqrt{t}\,y\big) - f\big(x + \sqrt{t}\,y\big)}{h}\right| = \left|\frac{\partial f}{\partial x_i}\big(\xi + \sqrt{t}\,y\big)\right| \le M e^{\alpha\|\xi + \sqrt{t}\,y\|} \le M e^{\alpha(\|x\| + \sqrt{t}\|y\| + 1)}$$

Ker zadnji izraz kot funkcija spremenljivke y pripada $L^1(\gamma_n)$, res smemo uporabiti izrek o dominirani konvergenci, torej parcialni odvod lahko nesemo pod integral. To pa smo hoteli dokazati. ■

Lema B.12.7. *Naj bo $f \in \mathcal{C}_\kappa(\mathbb{R}^n)$ in $t \ge 0$. Tedaj za vsak $\alpha \in \mathbb{R}$ velja:*

$$\lim_{\|x\|\to\infty} e^{\alpha\|x\|}(P_t f)(x) = 0$$

Dokaz. Za $t = 0$ je lema očitno pravilna, saj je $P_0 f = f \in \mathcal{C}_\kappa(\mathbb{R}^n)$. Naj bo zdaj $t > 0$. Ker ima f kompakten nosilec, obstaja tak $R > 0$, da za vsak $x \in \mathbb{R}^n$, za katerega je $\|x\| > R$, velja $f(x) = 0$. Naj bo $M := \sup_{x\in\mathbb{R}^n}|f(x)|$. Tedaj za $\|x\| > R$ velja:

$$\begin{aligned}\big|(P_t f)(x)\big| \;&\le \int \Big|f\big(x + \sqrt{t}\,y\big)\Big|\;\gamma_n(dy) = \int_{\|y\|>t^{-1/2}(\|x\|-R)} \Big|f\big(x + \sqrt{t}\,y\big)\Big|\;\gamma_n(dy) \le \\ &\le M \int_{\|y\|>t^{-1/2}(\|x\|-R)} \gamma_n(dy)\end{aligned}$$

Sledi:

$$\begin{aligned}e^{\alpha\|x\|}\,\big|(P_t f)(x)\big| \;&\le M\left(e^{\alpha(\|x\|-R)}e^{\alpha R}\right)\int_{\|y\|>t^{-1/2}(\|x\|-R)} \gamma_n(dy) \le \\ &\le M e^{\alpha R}\int_{\|y\|>t^{-1/2}(\|x\|-R)} e^{\alpha\sqrt{t}\|y\|}\gamma_n(dy)\end{aligned}$$

Ker je integrand v $L^1(\gamma_n)$, gre integral proti 0, ko gre x proti neskončno. To pa je bilo treba dokazati. ■

Naj bo zdaj $f \in \mathcal{C}_\kappa^2(\mathbb{R}^n)$. Naj bo ∂ kak parcialni odvod drugega reda. Torej je $\partial f \in \mathcal{C}_\kappa(\mathbb{R}^n)$ in po lemi B.12.7 je $P_t\partial f \in \mathcal{C}_0(\mathbb{R}^n)$. Toda po lemi B.12.6 je $P_t\partial f = \partial P_t f$. Od tod pa sledi, da je $P_t f \in D$. Operator G torej prostor $\mathcal{C}_\kappa^2(\mathbb{R}^n)$, ki je gost v $\mathcal{C}_0(\mathbb{R}^n)$, preslika v D. Iz izreka B.12.4 in trditve B.12.5 zdaj sledi, da je G res kar zaprtje operatorja $\frac{1}{2}\Delta$, definiranega na D.

4. Feller–Dynkinove difuzije. Difuzija ne more kar tako skočiti čisto nekam drugam, ampak morajo biti njene trajektorije zvezne. Ogledali si bomo difuzije na $\mathbb{R}^n$, čeprav se da teorija posplošiti na poljubne končnorazsežne gladke mnogoterosti.

Definicija. *Feller–Dynkinova difuzija* na prostoru $S := \mathbb{R}^n$ je Feller–Dynkinov proces q na kompaktifikaciji $\widehat{S} := S \cup \{\infty\}$ prostora S, ki ima točko ∞ za absorbirajoče stanje in ima še naslednji dve lastnosti:

(1) Za vsak slučajni proces X z jedrom q je trajektorija $t \mapsto X_t(\omega)$ za skoraj vsak ω zvezna na množici $\{t \mid X_t(\omega) \neq \infty\}$.

(2) Definicijsko območje generatorja procesa vsebuje $\mathcal{C}_\kappa^\infty(\mathbb{R}^n)$, t. j. gladke funkcije s kompaktnim nosilcem.

Difuzija lahko torej skoči le v točko neskončno, drugje pa morajo biti trajektorije lepo zvezne.

Feller–Dynkinove difuzije se dajo zelo lepo karakterizirati. Njihov študij se da prevesti na študij eliptičnih diferencialnih operatorjev drugega reda. Velja namreč naslednji izrek.

Izrek B.12.8 (Dynkin, Kinney). *Naj bo q Feller–Dynkinov proces na $\widehat{S}$ in naj bo ∞ absorbirajoče stanje. Procesu potem pripada Feller–Dynkinova polgrupa na S z generatorjem G. Naslednji trditvi sta ekvivalentni:*

(1) *Proces je Feller–Dynkinova difuzija.*

(2) *Zožitev generatorja G na prostor $\mathcal{C}_\kappa^\infty(S)$ je eliptični diferencialni operator oblike:*

$$(Gf)(x) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x)\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x) + \sum_{i=1}^{n} b_i(x)\frac{\partial f}{\partial x_i}(x) - c(x)\,f(x) \qquad \text{(B.12.6)}$$

kjer so funkcije a_{ij}, b_i in c zvezne in za vsak x velja, da je matrika $[a_{ij}(x)]_{i,j}$ pozitivno semidefinitna in $c(x) \geq 0$.

Dokaz bomo opustili, bralec ga najde v [21] na straneh 258–260. Iz izreka sledi tudi izrek B.10.1. Najprej dobimo, da ima Brownovo gibanje zvezne trajektorije, in tako je izrek v eno smer že dokazan. Dokaz v obratno smer pa najde bralec v [21] na straneh 261–262.

Opomba. Zveznost trajektorij v izreku B.10.1 je potreben pogoj za enoličnost. V nasprotnem primeru namreč lahko vzamemo simetrični slučajni sprehod na $\mathbb{Z}$ v zveznem času, ki ga s translacijami razširimo na $\mathbb{R}$. Generator tega procesa je:

$$(Gf)(x) := \lambda\big(f(x+1) - 2f(x) + f(x-1)\big)$$

Ni težko preveriti, da so lokalna jedra tega procesa podana po predpisu:

$$q_t(x, A) := \mathbf{P}[x + U_t - V_t \in A]$$

kjer sta U_t in V_t neodvisni slučajni spremenljivki, porazdeljeni po $\mathrm{Po}(\lambda t)$, torej imamo dejansko proces na $\mathbb{R}$ in ne na $\mathbb{R} \cup \{\infty\}$. Podoben proces lahko konstruiramo tudi na $\mathbb{R}^n$, tako da po komponentah sestavimo neodvisne procese na $\mathbb{R}$.

Nastane seveda vprašanje, kdaj operator G, definiran po formuli (B.12.6), res določa Feller–Dynkinov proces. Razmeroma enostavno se da preveriti, da je operator G disipativen. Po lemi 2.11 na strani 16 v [10] se da G zapreti in velja $\mathcal{R}(\lambda I - \overline{G}) = \overline{\mathcal{R}(\lambda I - G)}$. Treba je preveriti, da je to res cel prostor, torej rešiti parcialno diferencialno enačbo. Le-to pa si lahko bralec pogleda v [22], razdelek V.22. Podoben problem, le da je S kompaktno območje v $\mathbb{R}^n$ z robom iz razreda $\mathcal{C}^2$, je obdelan v [10] na straneh 366–369 in [15] (izrek 3.13).

B.13 Ornstein–Uhlenbeckov proces

Brownovo gibanje ima kot model slučajnega gibanja molekul bistveno fizikalno pomanjkljivost: njegove trajektorije so sicer skoraj gotovo zvezne, vendar pa se izkaže, da tudi skoraj gotovo niso nikjer odvedljive. Se pa da iz Brownovega gibanja izpeljati tudi tak model, in sicer tako, da namesto položaja gledamo *hitrost*. Tak model je lepo usklajen z Newtonovimi zakoni: impulzi sile so porazdeljeni povsem slučajno, tako kot naše matematično Brownovo gibanje.

V takem modelu lahko hitrost precej podivja. V resnici pa se molekuli, ki se giblje prehitro, hitrost naglo zmanjša. Zato v naš model vgradimo še "trenje". Na hitrost torej vplivajo slučajni impulzi sile, ki so porazdeljeni tako kot Brownovo gibanje, poleg tega pa nanjo vpliva še deterministična sila, ki je nasprotno sorazmerna s hitrostjo molekule.

Naj bo torej u trajektorija Brownovega gibanja. Privzemimo za hip, da je $u \in \mathcal{C}^1([0,\infty), \mathbb{R}^n)$. Iz nje moramo narediti trajektorijo v, ki bo kazala hitrost molekule. Vektorska funkcija v zadošča diferencialni enačbi:

$$\dot{v} = \dot{u} - cv, \quad v(0) = x \tag{B.13.1}$$

Parameter x je začetna hitrost molekule. Rešitev te diferencialne enačbe je:

$$v(t) = e^{-ct}\left(x + \int_0^t \dot{u}(s)e^{cs}\,ds\right) \tag{B.13.2}$$

Kot smo že omenili, pa trajektorije Brownovega gibanja niti slučajno niso zvezno odvedljive. Toda namesto sile $\dot{u}$ lahko gledamo tudi impulz sile u. Vrednost $u(t)$ naj bo torej skupen impulz sile, ki ga prejme naša molekula od časa 0 do t. Začetni impulz sile, torej vrednost $u(0)$, pa bo začetna hitrost naše molekule x. Porazdelitev slučajne funkcije u bo ravno Brownovo gibanje. Začetna porazdelitev tega Brownovega gibanja pa bo enaka porazdelitvi začetne hitrosti x.

Da se znebimo odvodov, zapišemo prejšnjo diferencialno enačbo v integralski obliki. Dobimo:

$$v(t) = u(t) - c\int_0^t v(s)\,ds \tag{B.13.3}$$

Iz enačbe jasno sledi, da je $v(0) = u(0) = x$, torej je $u(0)$ res začetna hitrost. Dobljena enačba je v primeru, ko je u zvezno odvedljiva, ekvivalentna stari diferencialni enačbi skupaj z začetnim pogojem. Toda nova enačba je zdaj definirana za poljubno zvezno preslikavo u. Treba jo je le še rešiti. Rešitev (B.13.2) ni dobra, saj mora biti u vsaj skoraj povsod odvedljiva, odvod pa integrabilen, če želimo, da je desna stran formule (B.13.2) dobro definirana. Tega pa se znebimo tako, da formulo (B.13.2) integriramo per partes. Dobimo:

$$v(t) = u(t) - c\,e^{-ct}\int_0^t u(s)\,e^{cs}\,ds$$

Krajši račun pokaže, da je tako definirana funkcija v res rešitev prejšnje integralske enačbe. To bo torej iskana hitrost molekule. Definirajmo torej funkcijo F: $\mathcal{C}([0,\infty), \mathbb{R}^n) \to \mathcal{C}([0,\infty), \mathbb{R}^n)$ po predpisu:

$$F(u)(t) := u(t) - c\,e^{-ct}\int_0^t u(s)\,e^{cs}\,ds$$

Zdaj pa moramo konstruirati še markovski proces, katerega trajektorije bodo tako transformirane trajektorije Brownovega gibanja. Natančneje, konstruirati bo treba markovsko jedro, torej transformacijo $\nu \mapsto \mathbf{P}_\nu$, da bo veljalo: Če je B Brownovo gibanje z začetno porazdelitvijo ν, ima $F(B)$ porazdelitev $\mathbf{P}_\nu$.

Brownovo gibanje lahko opišemo z markovskim jedrom ali pa s transformacijo $\nu \mapsto \mathbf{W}_\nu$ (torej je $\mathbf{W}_0 = \mathbf{W}$). Novo markovsko jedro pa bomo opisali s transformacijo $\nu \mapsto \mathbf{P}_\nu$, kjer bo:

$$\mathbf{P}_\nu := F_* \mathbf{W}_\nu$$

Če smo zelo dlakocepski, moramo v definiciji porazdelitev $\mathbf{W}_\nu$ najprej zožiti na množico vseh zveznih funkcij. Ta zožitev seveda ne spremeni bistveno porazdelitve $\mathbf{W}_\nu$, saj so po izreku B.12.8 trajektorije Brownovega gibanja skoraj gotovo zvezne.

Definirali smo torej transformacijo $\nu \mapsto \mathbf{P}_\nu$, zdaj pa je treba še preveriti, da nam res določa markovsko jedro. Naj bo p projekcija, ki preslikavi u priredi vrednost $u(0)$. Očitno je $p \circ F = p$, torej je tudi $p_* \mathbf{P}_\nu = p_* F_* \mathbf{W}_\nu = p_* \mathbf{W}_\nu = \nu$. Dobljena transformacija je torej desni inverz projekcije p. Velja tudi:

$$\mathbf{P}_\nu(A) = \mathbf{W}_\nu(F^{-1}(A)) = \int \mathbf{W}_x(F^{-1}(A))\, \nu(dx) = \int \mathbf{P}_x(A)\, \nu(dx)$$

torej je transformacija $\nu \mapsto \mathbf{P}_\nu$ usklajena tudi z integracijo, zato po izreku B.4.6 določa jedro q, za katerega pa je treba še preveriti, da je markovsko. Naj bo torej V slučajni proces z jedrom q. Preveriti moramo, da je za vsak $t \in [0, \infty)$ slučajna spremenljivka $\theta_t(V)$ pogojno glede na V_t neodvisna od $\mathcal{F}_t = \sigma(V_s \mid 0 \le s \le t)$ in da je tudi sama slučajni proces z jedrom q.

Ker je V slučajni proces z jedrom q, ima porazdelitev $\mathbf{P}_\nu$ za neko začetno porazdelitev ν. To pa je porazdelitev slučajne spremenljivke $F(B)$, kjer je B Brownovo gibanje z začetno porazdelitvijo ν (in s trajektorijami, ki so čisto gotovo, ne le skoraj gotovo zvezne). Brez škode za splošnost lahko torej privzamemo, da je kar $V = F(B)$, saj je to, ali je dana slučajna spremenljivka markovski proces ali slučajni proces z jedrom q, odvisno le od njene porazdelitve.

Naj bo $u \in \mathcal{C}([0, \infty), \mathbb{R}^n)$. Funkcija $v := F(u)$ je rešitev integralske enačbe (B.13.3). To je tudi edina rešitev te integralske enačbe. Če bi namreč obstajala še ena rešitev v', bi razlika $w := v' - v$ zadoščala integralski enačbi:

$$w(t) = -c \int_0^t w(s)\, ds$$

Od tod pa sledi, da je funkcija w zvezno odvedljiva, potem pa enačbo lahko odvajamo in spet dobimo diferencialno enačbo $\dot{w} = -cw$ z začetnim pogojem $w(0) = 0$. Le-ta pa ima edino rešitev $w = 0$.

Za funkcijo $\theta_t(v)$ pa velja:

$$\begin{aligned}
\theta_t(v)(s) = v(t+s) = u(t+s) - c \int_0^{t+s} v(r)\, dr = \\
= u(t+s) - c \int_0^t v(r)\, dr - c \int_t^{t+s} v(r)\, dr = \\
= \theta_t(u)(s) - u(t) + v(t) - c \int_0^s \theta_t(v)(r)\, dr
\end{aligned}$$

Funkcija $\theta_t(v)$ torej reši integralsko enačbo (B.13.3), pri čemer u zamenjamo s $\theta_t(u) - u(t) + F(u)(t)$. Iz enoličnosti rešitve sledi:

$$\theta_t(F(u)) = F\big(\theta_t(u) - u(t) + F(u)(t)\big)$$

oziroma:

$$\theta_t(V) = F\big(\theta_t(B) - B_t + V_t\big)$$

Po izreku B.9.1 je izraz pod F spet Brownovo gibanje z neko začetno porazdelitvijo ν, torej ima $\theta_t(X)$ porazdelitev $\mathbf{P}_\nu$. Od tod pa sledi, da je $\theta_t(X)$ slučajni proces z jedrom q. Preveriti moramo le še, da je pogojno glede na V_t neodvisen od $\mathcal{F}_t$.

Naj bo $\mathcal{D}_t := \sigma(\theta_t(B) - B_t)$. Po izreku B.9.1 je ta σ-algebra neodvisna od σ-algebre, ki jo generirajo slučajne spremenljivke B_s, $0 \le s \le t$, le-ta pa vsebuje σ-algebro $\mathcal{F}_t$ (to takoj sledi iz definicije funkcije F). Slučajna spremenljivka $\theta_t(V)$ je merljiva glede na σ-algebro $\mathcal{E}_t := \sigma(\mathcal{D}_t, V_t)$, torej bo dovolj pokazati, da je σ-algebra $\mathcal{E}_t$ pogojno glede na V_t neodvisna od $\mathcal{F}_t$.

Množica dogodkov iz $\mathcal{E}_t$, ki so pogojno glede na V_t neodvisni od $\mathcal{F}_t$, je λ-sistem. Po Dynkinovi lemi je dovolj dokazati, da vsebuje kak π-sistem, ki generira $\mathcal{E}_t$. Ta π-sistem pa bodo dogodki oblike $A \cap B$, kjer je $A \in \mathcal{D}_t$ in $B \in \sigma(V_t)$. Ker velja:

$$\begin{aligned}\mathbf{P}\big(A \cap B \,\big|\, \sigma(V_t)\big) &= \mathbf{E}\big(I_A I_B \,\big|\, \sigma(V_t)\big) = I_B \mathbf{E}\big(I_A \,\big|\, \sigma(V_t)\big) = I_B \mathbf{P}(A) = \\ &= I_B \mathbf{E}\big(I_A \,\big|\, \mathcal{F}_t\big) = \mathbf{E}\big(I_A I_B \,\big|\, \mathcal{F}_t\big) = \mathbf{P}\big(A \cap B \,\big|\, \mathcal{F}_t\big)\end{aligned}$$

je po izreku B.5.2 dogodek $A \cap B$ pogojno glede na V_t res neodvisen od $\mathcal{F}_t$. Potem pa to velja tudi za $\theta_t(V)$. To pa pomeni, da je jedro q res markovsko. Imenujemo ga *Ornstein–Uhlenbeckov proces*.

Izračunajmo operatorsko polgrupo tega procesa! Najprej bomo izračunali lokalna jedra q_t. Če je V slučajni proces z jedrom q, je q_t prehodna verjetnost slučajne spremenljivke V_t glede na V_0. Če je V_0 skoraj gotovo enaka x, je po formuli (B.1.1) kar $q_t(x, A) = \mathbf{P}[V_t \in A]$. Funkcija $A \mapsto q_t(x, A)$ je torej porazdelitev slučajne spremenljivke V_t.

Naj bo B Brownovo gibanje z začetkom v x in naj bo $V := F(B)$ (dlakocepski bralec bo spet pripomnil, da morajo biti trajektorije čisto gotovo, ne le skoraj gotovo zvezne). Tedaj V zadošča zahtevam prejšnjega odstavka. Če definiramo:

$$F_t(u) := F(u)(t) = u(t) - c\, e^{-ct} \int_0^t u(s)\, e^{cs}\, ds$$

je $V_t = F_t(B)$. Porazdelitev te slučajne spremenljivke bomo izračunali tako, da jo bomo zapisali kot limito slučajnih spremenljivk, pri kateri bomo integral zamenjali z Riemannovo vsoto. Nato se bomo oprli na izrek A.2.5. Definirajmo torej funkcije $F_t^{(n)}$: $\mathcal{C}([0,\infty), \mathbb{R}^n) \to \mathbb{R}$ po predpisu:

$$F_t^{(n)}(u) := u(t) - \frac{ct}{n}\, e^{-ct} \sum_{k=0}^{n-1} u\big(\tfrac{kt}{n}\big)\, e^{c\frac{kt}{n}}$$

Tedaj za vsak u zaporedje $F_t^{(n)}(u)$ konvergira proti $F_t(u)$. Ugotoviti moramo porazdelitev slučajnih spremenljivk:

$$F_t^{(n)}(B) = B_0 - \frac{ct}{n}\, e^{-ct} \sum_{k=0}^{n-1} B_{\frac{kt}{n}}\, e^{c\frac{kt}{n}}$$

ki po točkah konvergirajo proti V_t. Definirajmo slučajne spremenljivke $Z_1^{(n)}, \dots Z_n^{(n)}$ po predpisu:

$$Z_i^{(n)} := B_{\frac{ti}{n}} - B_{\frac{t(i-1)}{n}}$$

Po posledici B.9.2 in formuli (B.10.2) so slučajne spremenljivke $Z_1^{(n)}, \dots Z_n^{(n)}$ neodvisne in porazdeljene po $\mathrm{N}\big(0, \frac{t}{n} I\big)$. Ker za $k = 0, \dots n$ skoraj gotovo velja:

$$B_{\frac{kt}{n}} = x + \sum_{i=1}^{k} Z_i^{(n)}$$

tudi skoraj gotovo velja:

$$F_t^{(n)}(B) = x \left(1 - \frac{ct}{n} e^{-ct} \sum_{k=0}^{n-1} e^{c\frac{kt}{n}}\right) + \sum_{i=1}^{n} Z_i^{(n)} - \frac{ct}{n} e^{-ct} \sum_{k=0}^{n-1} e^{c\frac{kt}{n}} \sum_{i=1}^{k} Z_i^{(n)}$$

Slučajne spremenljivke $F_t^{(n)}(B)$ so torej porazdeljene normalno. Velja:

$$\mathbf{E}\big(F_t^{(n)}(B)\big) = x \left(1 - \frac{ct}{n}\, e^{-ct} \sum_{k=0}^{n-1} e^{c\frac{kt}{n}}\right)$$

Vsota v izrazu na desni je seveda ravno Riemannova vsota funkcije $s \mapsto e^{cs}$ na intervalu $[0, t]$. Sledi:

$$\lim_{n\to\infty} \mathbf{E}\big(F_t^{(n)}(B)\big) = x \left(1 - c\, e^{-ct} \int_0^t e^{cs}\, ds\right) = x\big(1 - (1 - e^{-ct})\big) = e^{-ct} x$$

Izračunajmo še varianco! Velja $\mathrm{var}\big(F_t^{(n)}(B)\big) = \mathrm{var}\big(W_t^{(n)}\big)$, kjer je:

$$\begin{aligned}
W_t^{(n)} &= \sum_{i=1}^{n} Z_i^{(n)} - \frac{ct}{n}\, e^{-ct} \sum_{k=0}^{n-1} e^{c\frac{kt}{n}} \sum_{i=1}^{k} Z_i^{(n)} = \sum_{i=1}^{n} Z_i^{(n)} - \frac{ct}{n}\, e^{-ct} \sum_{i=1}^{n-1} \sum_{k=i}^{n} e^{c\frac{kt}{n}} Z_i^{(n)} = \\
&= \sum_{i=1}^{n} Z_i^{(n)} - \frac{ct}{n}\, e^{-ct} \sum_{i=1}^{n-1} \frac{e^{ct} - e^{\frac{cit}{n}}}{e^{\frac{ct}{n}} - 1} Z_i^{(n)} = \sum_{i=1}^{n} \left(1 - \frac{ct}{n}\, e^{-ct} \frac{e^{ct} - e^{\frac{cit}{n}}}{e^{\frac{ct}{n}} - 1}\right) Z_i^{(n)} = \\
&= \sum_{i=1}^{n} \left(1 - \frac{ct}{n}\, \frac{1 - e^{-\frac{c(n-i)t}{n}}}{e^{\frac{ct}{n}} - 1}\right) Z_i^{(n)} = \sum_{i=0}^{n-1} \left(1 - \frac{ct}{n}\, \frac{1 - e^{-\frac{cit}{n}}}{e^{\frac{ct}{n}} - 1}\right) Z_{n-i}^{(n)}
\end{aligned}$$

Zdaj pa upoštevamo, da so slučajne spremenljivke $Z_1^{(n)}, \dots Z_n^{(n)}$ neodvisne in porazdeljene po $\mathrm{N}\big(0, \frac{t}{n} I\big)$. Od tod sledi, da je $\mathrm{var}\big(F_t^{(n)}(B)\big) = w_t^{(n)} I$, kjer je:

$$\begin{aligned}
w_t^{(n)} &= \sum_{i=0}^{n-1} \left(1 - \frac{ct}{n}\, \frac{1 - e^{-\frac{cit}{n}}}{e^{\frac{ct}{n}} - 1}\right)^2 \frac{t}{n} = \\
&= \sum_{i=0}^{n-1} \frac{t}{n} - 2 \frac{\frac{ct}{n}}{e^{\frac{ct}{n}} - 1} \sum_{i=0}^{n-1} \Big(1 - e^{-\frac{cit}{n}}\Big) \frac{t}{n} + \left(\frac{\frac{ct}{n}}{e^{\frac{ct}{n}} - 1}\right)^2 \sum_{i=0}^{n-1} \Big(1 - e^{-\frac{cit}{n}}\Big)^2 \frac{t}{n}
\end{aligned}$$

Pošljimo n proti neskončno! Izraz $\dfrac{\frac{ct}{n}}{e^{\frac{ct}{n}}-1}$ gre proti 1, vse vsote v zadnjem izrazu pa so Riemannove vsote primernih funkcij na intervalu $[0,t]$. Sledi:

$$\begin{aligned}\lim_{n\to\infty} w_t^{(n)} &= \int_0^t \Big(1-2\big(1-e^{-cs}\big)+\big(1-e^{-cs}\big)^2\Big)ds = \int_0^t \Big(1-\big(1-e^{-cs}\big)^2\Big)ds = \\ &= \int_0^t e^{-2cs}\,ds = \frac{1-e^{-2ct}}{2c}\end{aligned}$$

Iz posledice A.2.6 zdaj dobimo, da zaporedje porazdelitev slučajnih spremenljivk $F_t^{(n)}(B)$ šibko konvergira k porazdelitvi:

$$\mathrm{N}\left(e^{-ct}x, \frac{1-e^{-2ct}}{2c}I\right)$$

Iz trditve A.2.5 pa sledi, da zaporedje teh porazdelitev konvergira tudi k porazdelitvi slučajne spremenljivke $F_t(B)$. Ker je po trditvi A.2.3 prostor porazdelitev v šibki topologiji Hausdorffov, mora biti:

$$V_t = F_t(B) \sim \mathrm{N}\left(e^{-ct}x, \frac{1-e^{-2ct}}{2c}I\right)$$

To pa pomeni, da je:

$$q_t(x,A) = \mathrm{N}\left(e^{-ct}x, \frac{1-e^{-2ct}}{2c}I\right)(A)$$

od koder dobimo tudi operatorsko polgrupo, saj po posledici B.3.3 velja:

$$(P_tf)(x) = \int f\left(e^{-ct}x + \sqrt{\frac{1-e^{-2ct}}{2c}}\,y\right)\gamma_n(dy)$$

Konstanta c je bistven parameter Ornstein–Uhlenbeckovega procesa, saj od ene vrednosti te konstante k drugi ne moremo preiti kar z linearno transformacijo (niti prostora niti časa). Navadno pa vzamemo $c = \frac{1}{2}$, čas pa vzamemo dvakrat hitrejši (t zamenjamo z $2t$). Dobimo:

$$(P_tf)(x) = \int f\left(e^{-t}x + \sqrt{1-e^{-2t}}\,y\right)\gamma_n(dy) \tag{B.13.4}$$

Zgornji operatorji seveda tvorijo krepko zvezno operatorsko polgrupo, saj pripadajo konkretnemu Feller–Dynkinovemu procesu. Ključen za krepko zveznost je izrek B.8.4. Seveda pa bi se dala krepka zveznost preveriti tudi neposredno, s tehniko epsilon–delta.

Izračunajmo generator Ornstein–Uhlenbeckovega procesa! To bomo storili podobno kot pri izpeljavi generatorja Brownovega gibanja – z razvojem v Taylorjevo vrsto. Naj bo torej $f \in \mathcal{C}_0(\mathbb{R}^n) \cap \mathcal{C}^2(\mathbb{R}^n)$ in naj bodo vsi drugi parcialni odvodi funkcije f v $\mathcal{C}_0(\mathbb{R}^n)$ (funkcija f je torej v definicijskem območju generatorja Brownovega gibanja). Razvoj v Taylorjevo vrsto nam podobno kot pri izpeljavi generatorja Brownovega gibanja da:

$$\begin{aligned}f\left(e^{-t}x + \sqrt{1-e^{-2t}}\,y\right) &= f(e^{-t}x) + \sqrt{1-e^{-2t}}\sum_{i=1}^n \frac{\partial f}{\partial x_i}(e^{-t}x)\,y_i + \\ &+ \frac{1-e^{-2t}}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f}{\partial x_i\,\partial x_j}(e^{-t}x)\,y_i\,y_j + R(x,y,t)\end{aligned}$$

kjer je:

$$|R(x,y,t)| \le \frac{1-e^{-2t}}{2}\varphi\big(\sqrt{1-e^{-2t}}\|y\|\big)\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i y_j|$$
$$\varphi(r) := \sup_{\substack{\|x-y\|\le r\\ 1\le i,j\le n}}\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(x) - \frac{\partial^2 f}{\partial x_i \partial x_j}(y)\right|$$

Ko integriramo po $\gamma_n(dy)$, dobimo:

$$(P_t f)(x) = f(e^{-t}x) + \frac{1-e^{-2t}}{2}(\Delta f)(e^{-t}x) + \int R(x,y,t)\,\gamma_n(dy)$$

Zgornji izraz lahko gledamo tudi kot preslikavo iz $[0,\infty)$ v $\mathcal{C}_0(\mathbb{R}^n)$ in izračunati moramo desni odvod te preslikave pri $t = 0$. Pišimo:

$$P_t f = u_1(t) + \frac{1-e^{-2t}}{2t}\big(u_2(t) + u_3(t) + u_4(t)\big)$$

kjer je:

$$u_1(t)(x) := f(e^{-t}x), \quad u_2(t)(x) := t(\Delta f)(x), \quad u_3(t)(x) := t\big((\Delta f)(e^{-t}x) - (\Delta f)(x)\big),$$
$$u_4(t)(x) := \frac{2t}{1-e^{-2t}}\int R(x,y,t)\,\gamma_n(dy), \quad u_4(0) := 0$$

Dovolj je preveriti, da so vse štiri funkcije z desne odvedljive v 0. Tudi pri odvajanju preslikav z vrednostmi v Banachovem prostoru namreč velja pravilo za odvajanje produkta: če je α odvedljiva realna funkcija, u pa odvedljiva funkcija z vrednostmi v Banachovem prostoru, je odvedljiva tudi funkcija αu in velja $(\alpha u)' = \alpha' u + \alpha u'$. V našem primeru je:

$$\alpha(t) = \begin{cases} \dfrac{1-e^{-2t}}{2t} & t \ne 0 \\ 1 & t = 0 \end{cases}$$

Funkcija α je odvedljiva (saj je celo analitična). Ker je $u_2(0) = u_3(0) = u_4(0) = 0$, velja:

$$Gf = \left(\frac{d}{dt}\right)^{+}\Bigg|_{t=0} P_t f = u_1'(0) + u_2'(0) + u_3'(0) + u_4'(0)$$

brž ko imajo vse preslikave u_1, u_2, u_3 in u_4 desni odvod v 0 (seveda v prostoru $\mathcal{C}_0(\mathbb{R}^n)$ s supremum normo). Za u_1 in u_2 bomo celo pokazali, da imata obojestranski odvod, kar bi se dalo pokazati tudi za u_3, preslikava u_4 pa za negativne t ni definirana. Če zapišemo:

$$u_1(t) = v(e^{-t}), \quad v(s)(x) := f(sx)$$

po verižnem pravilu velja $u_1'(0) = -v'(1)$. Velja:

$$v(1+h)(x) = f((1+h)x) = f(x) + h\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}((1+\vartheta h)x)\,x_i =$$
$$= v(1) + hx\cdot(\nabla f)((1+\vartheta h)x)$$

kjer je $0 \le \vartheta \le 1$. Domnevamo torej, da je:

$$v'(1)(x) = x \cdot (\nabla f)(x) \tag{B.13.5}$$

Ta funkcija pa bo morala biti v $\mathcal{C}_0(\mathbb{R}^n)$. Za funkcijo f bomo torej privzeli še, da je $\lim_{\|x\|\to\infty} x \cdot (\nabla f)(x) = 0$. Dokažimo zdaj, da je v res odvedljiva v 1. Naj bo $\varepsilon > 0$. Obstaja tak $r > 0$, da za vsak $y \in \mathbb{R}^n$, za katerega je $\|y\| > r$, velja $|y \cdot (\nabla f)(y)| < \varepsilon/3$. Naj bo zdaj $\|x\| > 2r$ in $|h| < \frac{1}{2}$. Tedaj je gotovo $\|(1+\vartheta h)x\| > r$. Sledi:

$$\begin{aligned}\left|\frac{v(1+h)(x)-v(1)(x)}{h} - x\cdot(\nabla f)(x)\right| &= \big|x\cdot(\nabla f)((1+\vartheta h)x) - x\cdot(\nabla f)(x)\big| \le \\ &\le \frac{1}{1+\vartheta h}|(1+\vartheta h)x\cdot(\nabla f)((1+\vartheta h)x)| + |x\cdot(\nabla f)(x)| < \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon\end{aligned}$$

Na množici $\{y \mid \|y\| \le 3r\}$ pa je preslikava ∇f enakomerno zvezna. Obstaja torej tak $\delta > 0$, da za poljubna y in z z normo največ $3r$ in oddaljena za manj kot δ velja $\|(\nabla f)(y) - (\nabla f)(z)\| < \frac{\varepsilon}{2r}$. Naj bo zdaj $\|x\| \le 2r$ in $|h| < \min\left\{\frac{1}{2}, \frac{\delta}{2r}\right\}$. Tedaj je gotovo $\|(1+\vartheta h)x\| \le 3r$ in $\|\vartheta h x\| < \delta$. Sledi:

$$\begin{aligned}\left|\frac{v(1+h)(x)-v(1)(x)}{h} - x\cdot(\nabla f)(x)\right| &= \big|x\cdot(\nabla f)((1+\vartheta h)x) - x\cdot(\nabla f)(x)\big| \le \\ &\le \|x\|\,\|(\nabla f)((1+\vartheta h)x) - (\nabla f)(x)\| < 2r\,\frac{\varepsilon}{2r} = \varepsilon\end{aligned}$$

Dokazali smo, da, brž ko je $|h| < \min\left\{\frac{1}{2}, \frac{\delta}{2r}\right\}$, velja:

$$\left\|\frac{v(1+h)-v(1)}{h} - \big(x \mapsto x\cdot(\nabla f)(x)\big)\right\| < \varepsilon$$

To pa pomeni, da je preslikava v odvedljiva v 1 in da velja (B.13.5), torej:

$$u_1'(0)(x) = -x\cdot(\nabla f)(x)$$

Očitno je $u_2'(0) = \Delta f$. Pokažimo še, da je $u_3'(0) = u_4'(0) = 0$, kjer pa tokrat mislimo desna odvoda. Ker je $u_3(0) = u_4(0) = 0$, bo dovolj pokazati, da je $\lim_{t\downarrow 0}\frac{\|u_3(t)\|}{t} = \lim_{t\downarrow 0}\frac{\|u_4(t)\|}{t} = 0$. Najprej je:

$$\frac{|u_3(t)(x)|}{t} = |(\Delta f)(e^{-t}x) - (\Delta f)(x)|$$

Naj bo $\varepsilon > 0$. Ker vsi drugi parcialni odvodi funkcije f pripadajo $\mathcal{C}_0(\mathbb{R}^n)$, obstaja tak $r > 0$, da za vsak $y \in \mathbb{R}^n$, za katerega je $\|y\| > r$, velja $|(\Delta f)(y)| < \varepsilon/2$. Naj bo zdaj $\|x\| > er$ in $0 \le t < 1$. Tedaj je $\|e^{-t}x\| > r$, zato velja:

$$|(\Delta f)(e^{-t}x) - (\Delta f)(x)| \le |(\Delta f)(e^{-t}x)| + |(\Delta f)(x)| < \varepsilon$$

Na množici $\{y \mid \|y\| \le er\}$ pa je funkcija Δf enakomerno zvezna. Obstaja torej tak $\delta > 0$, da za poljubna y in z z normo največ er in oddaljena za manj kot δ velja $|(\Delta f)(y) - (\Delta f)(z)| < \varepsilon$. Brez škode za splošnost lahko privzamemo, da je $\delta < er$. Naj bo zdaj

$\|x\| \leq er$ in $0 \leq t < -\log\left(1 - \frac{\delta}{er}\right)$. Tedaj je gotovo $\|e^{-t}x\| \leq er$ in $\|(1 - e^{-t})x\| < \delta$. Sledi:

$$|(\Delta f)(e^{-t}x) - (\Delta f)(x)| < \varepsilon$$

Dokazali smo, da, brž ko je $0 \leq t < \min\left\{1, -\log\left(1 - \frac{\delta}{er}\right)\right\}$, velja $\frac{\|u_3(t)\|}{t} < \varepsilon$. To pa pomeni, da je res $u_3'(0) = 0$.

Dokažimo še, da je tudi $u_4'(0) = 0$. Velja:

$$\frac{\|u_4(t)\|}{t} = \frac{2}{1 - e^{-2t}} \sup_x \int R(x, y, t)\, \gamma_n(dy) \leq \int \varphi\big(\sqrt{1 - e^{-2t}}\|y\|\big) \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i y_j|\, \gamma_n(dy)$$

Dokazati moramo, da gre izraz na desni proti 0, ko gre t proti 0. Ker so drugi parcialni odvodi funkcije f omejeni, je omejena tudi funkcija φ. Ker je $\int \sum_{i,j} |y_i y_j|\, \gamma_n(dy) < \infty$, lahko uporabimo izrek o dominirani konvergenci, torej je dovolj dokazati, da za vsak $y \in \mathbb{R}^n$ velja:

$$\lim_{t \downarrow 0} \varphi\big(\sqrt{1 - e^{-2t}}\|y\|\big) = 0$$

To pa velja, ker so drugi parcialni odvodi enakomerno zvezni, saj so v $\mathcal{C}_0(\mathbb{R}^n)$. Torej je tudi $u_4'(0) = 0$. Če zdaj seštejemo vse odvode funkcij u_j, dobimo naslednji izrek.

Izrek B.13.1. *Naj bo $(P_t)_{t \in [0,\infty)}$ operatorska polgrupa, ki pripada Ornstein–Uhlenbeckovemu procesu, in naj bo:*

$$D := \left\{ f \in \mathcal{C}_0(\mathbb{R}^n) \cap \mathcal{C}^2(\mathbb{R}^n) \;\middle|\; (\forall i, j) \tfrac{\partial^2 f}{\partial x_i \partial x_j} \in \mathcal{C}_0(\mathbb{R}^n),\ \lim_{\|x\| \to \infty} x \cdot (\nabla f)(x) = 0 \right\} \tag{B.13.6}$$

Naj bo G generator operatorske polgrupe $(P_t)_{t \in [0,\infty)}$. Tedaj je $D \subset \mathcal{D}(G)$ in za vsak $f \in D$ velja:

$$(Gf)(x) = (\Delta f)(x) - x \cdot (\nabla f)(x) \tag{B.13.7}$$

Tako kot pri generatorju Brownovega gibanja se je tudi tu smiselno vprašati, ali je G zaprtje svoje zožitve na D. Odgovor je seveda pritrdilen. Spet se bomo oprli na trditev B.12.5 in pokazali, da operatorji P_t prostor $\mathcal{C}_\kappa^2$ dvakrat zvezno odvedljivih funkcij s kompaktnim nosilcem preslikajo v D.

Naj bo torej $f \in \mathcal{C}_\kappa^2(\mathbb{R}^n)$ in naj bo ∂ kak parcialni odvod prvega ali drugega reda. Torej je $\partial f \in \mathcal{C}_\kappa(\mathbb{R}^n)$ in po lemi B.12.7 velja:

$$\lim_{\|x\| \to \infty} e^{\|x\|} (P_t \partial f)(x) = 0 \tag{B.13.8}$$

Lema B.13.2. *Naj bodo P_t operatorji, definirani po formuli (B.13.4), delujejo pa naj na prostoru eksponentno omejenih funkcij na $\mathbb{R}^n$. Naj bo $r \in \mathbb{N}$, naj bo $f \in \mathcal{C}^r(\mathbb{R}^n)$ in naj bo funkcija f skupaj z vsemi svojimi parcialni odvodi do vključno reda r eksponentno omejena. Naj bo ∂ kak parcialni odvod reda r. Tedaj za vsak $t \in [0,\infty)$ element $P_t f$ pripada $\mathcal{C}^r(\mathbb{R}^n)$ in velja $\partial P_t f = e^{-rt} P_t \partial f$.*

Dokaz. Dovolj je dokazati za $r = 1$, saj lahko za višje r dokažemo z indukcijo. Naj bo $(v_t f)(x) := f(tx)$ in naj bo $(T_t)_{t \in [0,\infty)}$ operatorska polgrupa Brownovega gibanja (T_t

so torej operatorji P_t v formuli (B.12.3)). Tedaj velja $P_t = v_{e^{-t}} T_{1-e^{-2t}}$. Iz leme B.12.6 zdaj sledi:

$$\partial P_t = \partial v_{e^{-t}} T_{1-e^{-2t}} = e^{-t} v_{e^{-t}} \partial T_{1-e^{-2t}} = e^{-t} v_{e^{-t}} T_{1-e^{-2t}} \partial = e^{-t} P_t \partial$$

■

Iz formule (B.13.8), zgornje leme in ocene $1 \le \|x\| \le e^{\|x\|}$, zdaj dobimo:

$$\lim_{\|x\|\to\infty} (\partial P_t f)(x) = 0 \quad \text{in tudi} \quad \lim_{\|x\|\to\infty} \|x\| (\partial P_t f)(x) = 0$$

To pa pomeni, da je $P_t f \in D$. Operator G torej prostor $\mathcal{C}^2_\kappa(\mathbb{R}^n)$, ki je gost v $\mathcal{C}_0(\mathbb{R}^n)$, preslika v D. Iz izreka B.13.1 in trditve B.12.5 zdaj sledi, da je G res kar zaprtje operatorja, definiranega na D po formuli (B.13.7).

B.14 Ravnovesne porazdelitve

Dostikrat nas zanima, kdaj je dani slučajni proces X *stacionaren*, torej kdaj imajo vse slučajne spremenljivke X_t isto porazdelitev μ. Porazdelitvi μ bomo rekli *ravnovesna porazdelitev*. Pri Chen–Steinovi metodi pa nas bo zanimalo tudi, ali in kako hitro se porazdelitev danega slučajnega procesa bliža ravnovesni.

Definicija. Porazdelitev μ je *ravnovesna porazdelitev* markovskega jedra q, če je vsak slučajni proces z jedrom q in začetno porazdelitvijo μ stacionaren.

Trditev B.14.1. *Naj bo q markovsko jedro na prostoru stanj $(S, \mathcal{S})$ in s pripadajočo polgrupo $(P_t)_{t\in[0,\infty)}$ operatorjev na $\flat(S, \mathcal{S})$. Tedaj je μ ravnovesna porazdelitev procesa q natanko tedaj, ko za vsako funkcijo $f \in \flat(S, \mathcal{S})$ velja:*

$$\int P_t f \, d\mu = \int f \, d\mu$$

Dokaz. Naj bo μ ravnovesna porazdelitev procesa q. Naj bodo Y_t zožitve koordinatnih projekcij $S^{[0,\infty)} \to S$ na prostor $\overline{S}$. Glede na $\mathbf{P}_\mu$ je tedaj $(Y_t)_{t\in[0,\infty)}$ slučajni proces z začetno porazdelitvijo μ in jedrom q. Torej so jedra q_t, definirana po formuli (B.6.1), prehodne porazdelitve slučajnih spremenljivk Y_t glede na Y_0. Po posledici B.3.3 je potem za vsak t iz vsak $f \in \flat(S, \mathcal{S})$ funkcija $P_t f$ regresija slučajne spremenljivke $f(Y_t)$ glede na Y_0. Z drugimi besedami, velja $(P_t f)(Y_0) = \mathbf{E}_\mu(f(Y_t) \mid Y_0)$. Iz dejstva, da imata Y_0 in Y_t porazdelitev μ, potem sledi:

$$\int P_t f \, d\mu = \mathbf{E}_\mu\big((P_t f)(Y_0)\big) = \mathbf{E}_\mu\big(\mathbf{E}_\mu(f(Y_t) \mid Y_0)\big) = \mathbf{E}_\mu(f(Y_t)) = \int f \, d\mu$$

Dokažimo zdaj še obratno. Naj bo X slučajni proces z začetno porazdelitvijo μ in jedrom q. Tedaj je $P_t f$ regresija slučajne spremenljivke X_t glede na X_0 za vsako funkcijo $f \in \flat(S, \mathcal{S})$. Z drugimi besedami, velja $(P_t f)(X_0) = \mathbf{E}(f(X_t) \mid X_0)$. Sledi:

$$\mathbf{E}(f(X_t)) = \mathbf{E}\big(\mathbf{E}(f(X_t) \mid X_0)\big) = \mathbf{E}\big((P_t f)(Y_0)\big) = \int P_t f \, d\mu = \int f \, d\mu = \mathbf{E}(f(X_0))$$

To pa pomeni, da mora imeti X_t enako porazdelitev kot X_0, to pa je seveda μ. Slučajni proces X je torej res stacionaren. ■

Trditev B.14.2. *Naj bo q Feller–Dynkinov proces na lokalno kompaktnem Hausdorffovem prostoru stanj S s števno bazo. Procesu naj pripada ustrezna Feller–Dynkinova polgrupa $(P_t)_{t\in[0,\infty)}$ z generatorjem G. Naj bo μ porazdelitev na S. Naslednje trditve so ekvivalentne:*

(1) *μ je ravnovesna porazdelitev procesa q.*

(2) *Za vsak $t \in [0,\infty)$ in vsak $f \in \mathcal{C}(S)$ velja:*

$$\int P_t f \, d\mu = \int f \, d\mu$$

(3) *Za vsak $f \in \mathcal{D}(G)$ velja:*

$$\int Gf \, d\mu = 0$$

Dokaz. Implikacija (1) ⇒ (2) takoj sledi iz prejšnje trditve, pri implikaciji (2) ⇒ (1) pa je dovolj pokazati, da enačba velja za vsako funkcijo $f \in \flat(S, \mathcal{S})$, ne le za $f \in \mathcal{C}(S)$. Prostor S je normalen, saj je po Urisonovem metrizacijskem izreku celo metrizabilen. Ker ima števno bazo, lahko indikator vsake zaprte množice zapišemo kot limito padajočega zaporedja primernih Urisonovih funkcij. Enačba torej velja za vse indikatorje zaprtih funkcij. Družina vseh množic $A \in \mathcal{S}$, za katere enačba velja za $f = I_A$, je λ-sistem, ki vsebuje π-sistem zaprtih množic, le-ta pa generira $\mathcal{S}$. Po Dynkinovi lemi potem enačba velja za vse indikatorje množic iz $\mathcal{S}$. Potem pa velja tudi za vse enostavne funkcije in končno za vse funkcije iz $\flat(S, \mathcal{S})$.

Implikacija (2) ⇒ (3) je očitna. Za dokaz obratne implikacije pa se je treba malo poglobiti v dokaz Hille–Yosidovega izreka, natančneje, v konstrukcijo polgrupe $(P_t)_{t\in[0,\infty)}$ iz generatorja G. Velja namreč (glej [10], strani 12–13):

$$P_t f = \lim_{\lambda \downarrow 0} e^{\lambda t G(\lambda I - G)^{-1}} f$$

Izkaže se namreč, da je operator $G(\lambda I - G)^{-1}$ povsod definiran in omejen. Eksponentno funkcijo pa definiramo kot običajno – za vsak omejen linearni operator A definiramo:

$$e^A := \sum_{n=0}^{\infty} \frac{A^k}{k!}$$

Integral:

$$\int \left(\lambda t G(\lambda I - G)^{-1}\right)^n d\mu$$

je za $n = 0$ enak $\int f \, d\mu$, sicer pa 0. Od tod pa že sledi, da je tudi $\int P_t f \, d\mu = \int f \, d\mu$. ■

Nastane vprašanja, ali ravnovesna porazdelitev obstaja, in če že obstaja, ali je ena sama. In če je ena sama, ali za vsak slučajni proces X z jedrom q porazdelitev slučajne spremenljivke X_t šibko konvergira proti ravnovesni. V splošnem je odgovor pozitiven le na prvo vprašanje (glej [16], stran 10). Procesa z več invariantnimi porazdelitvami tudi ni težko konstruirati, saj je vsaka porazdelitev pri mirujočem procesu ravnovesna. Če hočemo doseči konvergenco proti ravnovesni porazdelitvi (takemu procesu pravimo *ergodičen*), moramo imeti dovolj močan pretok med vsemi stanji. Ergodičnosti na sploh ni lahko dokazovati, še zlasti, če imamo na voljo le generator. Lep zadosten pogoj za ergodičnost sistema medsebojno učinkujočih delcev je naveden v [16] na strani 31. No, pri Chen–Steinovi metodi pa nas bo zanimala ergodičnost Ornstein–Uhlenbeckovega procesa.

Oglejmo si nekaj zgledov invariantnih porazdelitev.

Zgled 1: Poissonov proces. Trajektorije Poissonovega procesa so naraščajoče stopničaste funkcije, zato ni kakega velikega upanja, da bi imel ravnvesno porazdelitev. Pa recimo, da je μ ravnovesna porazdelitev. Označimo $\mu_k := \mu(\{k\})$. Enačba $\int Gf\,d\mu = 0$ se potem zapiše takole:

$$\sum_{k=0}^{\infty} \lambda\big(f(k+1) - f(k)\big)\mu_k = 0$$

Ker je $\lambda > 0$, lahko ta parameter pokrajšamo. Enačbo pa lahko zapišemo tudi takole:

$$\mu_0 f(0) + \sum_{k=1}^{\infty}(\mu_k - \mu_{k-1})f(k) = 0$$

Naj bo $n \in \mathbb{N}$ ter naj bo $f(n) = 1$ in $f(k) = 0$ za $k \neq 0$. Za $n = 0$ najprej dobimo, da je $\mu_0 = 0$, za vse nadaljnje n pa z indukcijo dobimo, da mora biti tudi $\mu_n = 0$. Mera μ je torej ničelna, taka mera pa ni verjetnostna. Poissonov proces torej nima invariantne porazdelitve (ima pa invariantno mero, ki je kar mera, ki šteje).

Zgled 2. Proces, ki sili le naprej ali navzven, ne more imeti invariantne porazdelitve, nekaj ga mora vleči tudi nazaj. V tem zgledu bomo Poissonovemu procesu pristrigli peruti. Spomnimo se zgleda na strani 122. Osebki naj se torej rojevajo s konstantno pogostnostjo λ, pogostnost umiranja pa je sorazmerna z velikostjo populacije. Dokazali smo že, da generator takega procesa, ki je definiran po predpisu:

$$(Gf)(n) := \lambda\big(f(n+1) - f(n)\big) + n\big(f(n-1) - f(n)\big)$$

izpolnjuje pogoje Hille–Yosidovega izreka, torej določa operatorsko polgrupo, s tem pa Feller–Dynkinov proces. Ker je koeficient skoka nazaj v stanju nič enak nič, lahko ta proces gledamo le na $\mathbb{N}_0 \cup \{\infty\}$, torej imamo proces rojevanja in umiranja. V resnici proces ne uide v neskončnost, kar je intuitivno precej jasno, saj to ne velja niti za Poissonov proces, ki ima trajektorije "v povprečju" višje. Eksakten dokaz tega dejstva pa ni tako očiten, saj je treba stvar izpeljati iz samega generatorja. Lahko si npr. pomagamo z izrekom 12.2 na strani 256 v [21], ki nam pove, kako dolgo proces čaka v danem stanju in kam gre potem.

Ker zdaj velikost populacije ne more kar tako rasti čez vse meje, se lahko vprašamo, ali ima naš proces invariantno porazdelitev μ, ki bo skoncentrirana na $\mathbb{N}_0$. Spet naj bo $\mu_n := \mu(\{n\})$. Zdaj dobimo, da za vsako funkcijo $f \in \mathcal{D}(G)$, torej tudi za vsako funkcijo na $\mathbb{N}_0$ s kompaktnim nosilcem, velja:

$$\sum_{k=0}^{\infty} \lambda\big(f(k+1) - f(k)\big) + \sum_{k=1}^{\infty} k\big(f(k-1) - f(k)\big)\mu_k = 0$$

Naj bo spet $f(n) = 1$ in $f(k) = 0$ za $k \neq n$. Dobimo:

$$\begin{aligned} -\lambda\mu_0 + \mu_1 &= 0 \\ \lambda\mu_{n-1} - \lambda\mu_n + (n+1)\mu_{n+1} - n\mu_n &= 0, \quad n = 1, 2, \ldots \end{aligned}$$

Od tod po krajšem računu z indukcijo dokažemo, da mora veljati:

$$\mu_n = \frac{\lambda^n}{n!}\mu_0$$

Če želimo, da bo mera μ verjetnostna, mora veljati $\mu_0 = e^{-\lambda}$. Potem pa μ ni nič drugega kot Poissonova porazdelitev $\mathrm{Po}(\lambda)$. To je tudi edina invariantna porazdelitev tega procesa, seveda na $\mathbb{N}_0$. In v resnici je ta proces rojevanja in umiranja ozadje Poissonove aproksimacije po Chen–Steinovi metodi.

Iz dejstva, da je μ invariantna porazdelitev, sledi tudi, da naš proces ne uide v neskončnost. Po definiciji se to skoraj gotovo ne zgodi glede na $\mathbf{P}_\mu$. Ker pa je $\mu(\{n\}) > 0$ za vsak $n \in \mathbb{N}_0$, iz usklajenosti z integracijo, t. j. formule (B.4.3), sledi, da se to skoraj gotovo ne zgodi tudi glede na $\mathbf{P}_n$. Potem pa proces skoraj gotovo ne uide v neskončnost, ne glede na to, kakšna je začetna porazdelitev.

Zgled 3: Brownovo gibanje. Tudi Brownovo gibanje sili preveč navzven, da bi lahko imelo kako invariantno porazdelitev. Pa recimo, da temu ni tako, torej obstaja invariantna porazdelitev μ. Naj bo X Brownovo gibanje z začetno porazdelitvijo μ. Po izreku B.9.1 sta slučajni spremenljivki X_0 in $X_1 - X_0$ neodvisni.

Naj bo ϕ karakteristična funkcija porazdelitve μ. Iz definicije Brownovega gibanja je razvidno, da ima $X_1 - X_0$ standardizirano normalno porazdelitev, katere karakteristična funkcija je definirana po predpisu $\psi(y) := e^{-\frac{1}{2}\|y\|^2}$. Ker mora zaradi stacionarnosti imeti tudi $X_1 = X_0 + (X_1 - X_0)$ porazdelitev μ, po konvolucijskem izreku velja $\phi = \phi\psi$. Ker ψ ni nikjer enaka 0, mora biti $\phi = 0$, taka funkcija pa ni karakteristična funkcija porazdelitve. Brownovo gibanje torej res nima invariantne porazdelitve.

Zgled 4: Ornstein–Uhlenbeckov proces. Videli smo, da Brownovo gibanje nima invariantne porazdelitve. Toda pri Ornstein–Uhlenbeckovem procesu viskoznost preprečuje hitrosti molekule, da bi zrasla čez vse meje in hitrost mora ostati v mejah normale. Oglejmo si še enkrat operatorsko polgrupo:

$$(P_t f)(x) := \int f\left(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\right) \gamma_n(dy)$$

Fiksirajmo x in pošljimo t proti neskončno! Po izreku o dominirani konvergenci dobimo:

$$\lim_{t\to\infty} (P_t f)(x) = \int f \, d\gamma_n$$

Trditev B.14.3. *Dano naj bo markovsko jedro na $(S, \mathcal{S})$ z operatorsko polgrupo $(P_t)_{t\in[0,\infty)}$ in naj bo μ porazdelitev na $(S, \mathcal{S})$. Če za vsak $f \in \flat(S, \mathcal{S})$ in vsak $x \in S$ velja $\lim_{t\to\infty}(P_t f)(x) = \int f\, d\mu$, je μ edina invariantna porazdelitev tega markovskega jedra.*

Dokaz. Izberimo poljuben $x \in S$. Iz verige enakosti:

$$\int P_t f \, d\mu = \lim_{s\to\infty} (P_s P_t f)(x) = \lim_{s\to\infty} (P_{s+t} f)(x) = \lim_{s\to\infty} (P_s f)(x) = \int f \, d\mu$$

in trditve B.14.1 sledi, da je μ res invariantna porazdelitev. Pa recimo, da bi bila ν še kaka invariantna porazdelitev. Tedaj za vsak t velja $\int P_t f \, d\nu = \int f \, d\nu$. Če zdaj pošljemo t proti neskončno in uporabimo izrek o dominirani konvergenci, dobimo, da mora biti $\int f \, d\mu = \int f \, d\nu$ za vsak $f \in \flat(S, \mathcal{S})$, torej mora biti $\nu = \mu$. ■

Opomba. Trditev lahko formuliramo tudi za Feller–Dynkinove procese in testne funkcije iz razreda $\mathcal{C}_0(S)$.

Iz trditve zdaj takoj sledi, da je standardizirana normalna porazdelitev na $\mathbb{R}^n$, t. j. γ_n oz. $\mathrm{N}(0, I)$, edina invariantna porazdelitev Ornstein–Uhlenbeckovega procesa. Še več, proces celo na neki način konvergira k tej porazdelitvi, kar je tudi osnova za Chen–Steinovo metodo. Vendar pa pri njej potrebujemo močnejšo konvergenco, kot je samo konvergenca po točkah. Za ta namen pa moramo Banachov prostor, na katerem deluje operatorska polgrupa, nekoliko prilagoditi.

Literatura

[1] Richard Arratia, Larry Goldstein, Louis Gordon: *Poisson approximation and the Chen–Stein method*, Statistical Science **5** (1990), 403–434

[2] Richard Arratia, Larry Goldstein, Louis Gordon: *Two moments suffice for Poisson approximation: the Chen–Stein method*, Ann. Prob. **17** (1989), 9–25

[3] A. D. Barbour: *Stein's Method*, neobjavljen rokopis, 1995. Dosegljiv na `ftp:iamassi.unizh.ch/pub/Barbour`.

[4] A. D. Barbour: *Stein's method for diffusion approximations*, Prob. Th. Rel. Fields **84** (1990), 297–322

[5] A. D. Barbour, G. K. Eagleson: *Poisson approximation for some statistics based on exchangeable trials*, Advances in Appl. Prob. **15** (1983), 565–600

[6] A. D. Barbour, Lars Holst, Svante Janson: *Poisson Approximation*, Clarendon Press, Oxford, 1992

[7] L. Le Cam: An approximation theorem for the Poisson binomial distribution, Pacific J. Math. **10** (1960), 1181–1197

[8] Y. S. Chow, H. Teicher: *Probability Theory, Independence, Interchangeability, Martingales*, Springer Verlag, 1978

[9] Richard Durrett: *Probability – Theory and Examples*, Duxbury Press, 1996

[10] Stewart N. Ethier, Thomas G. Kurtz: *Markov Processes: Characterization and Convergence*, John Willey & Sons, 1986.

[11] Larry Goldstein, Gesine Reinert: *Stein's method and the zero-bias transformation with application to simple random sampling*, Ann. Appl. Prob. **7** (1997), 935–952

[12] Larry Goldstein, Yosef Rinott: *Multivariate normal approximations by Stein's method and size bias couplings*, J. Appl. Prob. **33** (1996), 1–17

[13] F. Götze: *On the rate of convergence in the multivariate CLT*, Ann. Prob. **19** (1991), 724–739

[14] Soo-Thong Ho, Louis H. Y. Chen: *An L_p bound for the remainder in a combinatorial central limit theorem*, Ann. Prob. **6** (1978), 231–249

[15] O. A. Ladyženskaja, N. N. Ural'ceva: *Linear and Quasilinear Partial Differential Equations*, Academic, New York, 1968

[16] Thomas M. Liggett: *Interacting Particle Systems*, Springer Verlag, 1985

[17] W.–L. Loh: *Stein's method and multinomial approximation*, Ann. Appl. Prob. **2** (1992), 536–554

[18] M. Luk: *Stein's method for the gamma distribution and related statistical applications*, doktorska disertacija, Univ. Southern California, 1994

[19] E. Peköz: *Stein's method for geometric approximation*, Tech. Report #225, Stats Dept, U. California, Berkeley, 1995

[20] Daniel Revuz, Marc Yor: *Continuous Martingales and Brownian Motion*, Springer Verlag, 1994

[21] L. C. G. Rogers, David Williams: *Diffusions, Markov Processes, and Martingales*, Volume 1: *Foundations*, John Willey & Sons, 1994.

[22] L. C. G. Rogers, David Williams: *Diffusions, Markov Processes, and Martingales*, Volume 2: *Itô Calculus*, John Willey & Sons, 1994.

[23] Walter Rudin: Real and Complex Analysis, McGraw Hill, 1987

[24] Charles Stein: *Approximate Computation of Expectations*, IMS, Hayward, Calif., 1986

[25] Aihua Xia: *On using the first difference in the Chen–Stein method*, Ann. Appl. Prob. **7** (1997), 899–916