UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 1. stopnja

Timotej Akrapovič

Test Kolmogorova in Smirnova

Delo diplomskega seminarja

Mentor: doc. dr. Jaka Smrekar

Ljubljana, 2012

Kazalo

Test Kolmogorova in Smirnova

Povzetek

Delo diplomskega seminarja vsebuje samozadostno izpeljavo statistike Kolmogorova in Smirnova. Kolmogorov in Smirnov statistični test je neparametričen test za testiranje hipoteze, če je slučajni vzorec porojen po neki vnaprej predpostavljeni zvezni porazdelitvi, kjer parametrv ne smemo ocenjevati iz vzorca. Diplomsko delo vsebuje dokaz izreka Glivenko-Cantelli. Predstavim tudi, kako pod določenimi pogoji vzorčno empirično porazdelitveno funkcijo obravnavam kot Poissonov proces. Iz česar eksaktno izpeljem izračun P-vrednosti testa. Na koncu dodam še algoritem v programskem jeziku R, ki izračuna P-vrednosti za vzorce do velikosti 300.

The Kolmogorov-Smirnov test

Abstract

This graduation thesis includes a self-contained derivation of Kolmogorov-Smirnov D_n statistics. The Kolmogorov-Smirnov statistical test is a nonparametric test for testing whether a random sample belongs to a pre-specified distribution function in which the parameters cannot be estimated from the data. The thesis also contains a proof of the Glivenko-Cantelli theorem. It is also shown how, under certain circumstances, the empirical distribution function of a random sample can be treated as a Poisson process. From this the P-values of the test are precisely calculated. The thesis concludes by also presenting an algorithm in R programming language that calculates P-values for samples as large as 300.

Math. Subj. Class. (2010): 62G10, 60J27, 60-04, 60G17

Ključne besede: testiranje hipotez, neparametrične metode, test Kolmogorova in Smirnova, Poissonov proces.

Keywords: hypothesis testing, nonparametric methods, Kolmogorov-Smirnov test, Poisson process.

1. Uvod

Tipičen problem v uporabni statistiki je preverjanje pripadnosti slučajnega vzorca, neki vnaprej predpostavljeni zvezni porazdelitvi F_0. Na tem mestu si zastavimo vprašanje, kakšno merilo naj izberemo za določanje kriterija, kdaj naš vzorec zares pripada slučajni spremenljivki X. Najpreprostejši za testiranje zgornjega vprašanja je test hi-kvadrat, ki pa zaradi slabosti določanja razredov iz zvezne slučajne spremenljivke ni optimalen. Ustreznejšo metodo je leta 1933 podal Kolmogorov [4], ki je bila kasneje tabelirana s strani Smirnova. Metoda temelji na uporabi zahtevnih programskih izračunov. V dobi, ko cene računanja eksponentno padajo, so metode, odvisne od računalniške zmogljivosti, vedno bolj privlačna izbira. Tu pride do izraza uporabnost testa Kolmogorova in Smirnova. V svojem diplomskem delu bom z zgledi in teorijo natančno predstavil izpeljavo izračuna omenjenega testa. Verodostojnost izpeljave bom podkrepil z vsemi pripadajočimi dokazi. Priložil bom tudi algoritem v programskem jeziku R, ki hitro in natančno izračuna P-vrednosti testa za vzorce do velikosti 300.

2. Test Kolmogorova in Smirnova

Naj bo $\vec{X} = (X_1, ...X_n)$ slučajni vektor n neodvisnih realizacij slučajne spremenljivke $X : \Omega \to R$, kjer ima X porazdelitveno funkcijo F. V praksi testiramo, če je naš vzorec porojen po neki vnaprej predpostavljeni porazdelitvi F_o.
Predpostavimo ničelno hipotezo:

$$H_0 : \forall x \in \mathbb{R}, F(x) = F_0(x),$$

proti alternativi

$$H_A : \exists x \in \mathbb{R}, F(x) \neq F_0(x).$$

Definicija 2.1. ***Statistična hipoteza*** z oznako ***H*** je domneva o porazdelitvi slučajne spremenljivke X na populaciji.

Običajno tu uporabljamo hi-kvadrat test. test Kolmogorova in Smirnova ima pred hi-kvadrat testom vsaj dve veliki prednosti:

(1) Uporaben je z majhnimi vzorci, kjer je veljavnost testa hi-kvadrat vprašljiva.
(2) Pogosto se izkaže, da je močnejši od testa hi-kvadrat za katerokoli velikost vzorca.

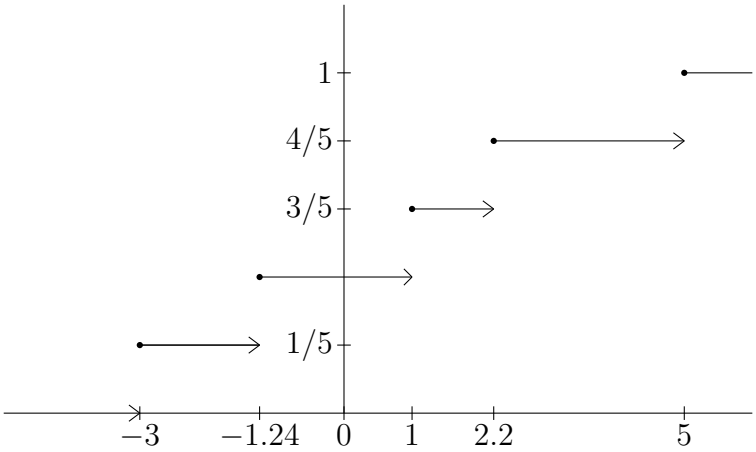
Test Kolmogorova in Smirnova meri ustreznost prileganja med empirično porazdelitveno funkcijo vzorca F_n in vnaprej predpostavljeno porazdelitveno funkcijo F_0 kot supremum razlik med F_n in F_0.

$$D_n = \sup_x \mid F_n(x) - F_0(x) \mid \tag{1}$$

Empirična porazdelitvena funkcija, ki ponazarja delež opazovanj manjših ali enakih vrednosti x, je definirana kot:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x} \tag{2}$$

Primer: Naj bo $s \in \Omega^5$ in $\vec{X}(s)$ z naslednjimi realizacijami $\vec{X}(s) = (X_1 = 3, X_2 = -1.24, X_3 = 1, X_4 = 2.2, X_5 = 5)$. Tedaj njegova empirična porazdelitvena funkcija izriše spodnji graf:



SLIKA 1. empirična porazdelitvena funkcija vzorca $\vec{X}(s)$

.

Verodostojnost testne statistike D_n zagovarjamo z ***Glivenko-Cantellijevim*** izrekom:

$$P(\lim_{n\to\infty} D_n = 0) = 1, \text{če je } F \equiv F_0 \tag{3}$$

Ta nam s povečevanjem velikosti vzorca preko vseh meja zagotavlja zavrnitev nepravilne ničelne hipoteze z verjetnostjo 1.
Pri testiranju hipotez se v splošnem sprašujemo, s kakšno verjetnostjo smo zavrnili pravilno ničelno hipotezo H_0, čemur pravimo napaka prve vrste. Zanima nas, pri kateri vrednosti testne statistike, v mojem primeru D_n, bomo ovrgli ničelno hipotezo. Ker je $D_n = \sup_x \mid F_n(x) - F_0(x) \mid$, pomeni, da se z večanjem vrednosti testne statistike pri velikosti vzorca n manjša verjetnost napake prve vrste. Zastavi se nam vprašanje o stopnji značilnosti pri zavrnitvi ničelne hipoteze, ki jo za neki vzorec velikosti n in testno statistiko D_n označimo z α. H_0 zavrnemo, ko je D_n element kritičnega območja $(d_\alpha, 1]$ pri vnaprej predpostavljeni stopnji značilnosti α. Centralna tema mojega diplomskega seminarja bo prav eksaktni izračun verjetnosti $P(D_n \leq d)$, s katero določimo najmanjšo stopnjo značilnosti, pri kateri lahko zavrnemo ničelno hipotezo oziroma P-vrednost. Ko imamo izračunano P-vrednost, H_0 zavrnemo, če je $\alpha >$ P-vrednost, sicer je ne zavrnemo.

$$\text{P-vrednost} = 1 - P(D_n \leq d).$$

3. Izrek Glivenka in Cantellija

Naj bo $F_n(x,s)$ zaporedje empiričnih porazdelitvenih funkcij, definirano kot:

$$F_n(x,s) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i(s)\leq x}, \quad s \in S = \Omega^n \tag{4}$$

Moj cilj je pokazati, da je $F_n(x,s)$ smiseln približek za $F(x)$, torej se z rastočim n za $\forall$ x ujema z $F(x)$. Sledi, $F_n(x,s)$ konvergira enakomerno proti $F(x)$. Spomnimo se na Krepki zakon velikih števil .

Izrek 3.1. (Kolmogorov krepki zakon velikih števil): *Če so slučajne spremenljivke $\{X_i\}_{i=1}^{\inf}$ neodvisne, enako porazdeljene in vse element L^1 ter je $S_n = \sum_{i=1}^{n} X_i$, potem velja: $\frac{S_n}{n} \to EX$ s.g.*

Označimo z $Y_i(s) = \mathbb{1}_{(-\infty,x](X_i(s))}$, torej je Y Bernulijeva slučajna spremenljivka, porazdeljena kot:

$$Y_i(s) \sim \begin{pmatrix} 1 & 0 \\ \text{F(x)} & \text{1 - F(x)} \end{pmatrix}$$

Uporabimo Krepki zakon velikih števil :

$$Y_i(s) \in L^1 \Rightarrow \lim_{n\to\infty} F_n(x,s) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i(s) \to EY s.g = F(x),$$

ki nam omogoča konvergenco $F_n(x,s) \to F(x)$ za $\forall x$. V resnici pa nam spodnji izrek poda še močnejšo lastnost.
Dokaz spodaj omenjenega izreka je bil podan v zapiskih predavanj pri predmetu Verjetnost in statistika.

Izrek 3.2. (Glivenko-Cantelli): $P(\lim_{n\to\infty} D_n = 0) = 1$.

Dokaz Glivenko-Cantellijevega izreka. Definiramo:

$$D_n(s) := \sup_{x\in\mathbb{R}} \mid F_n(x,s) - F(x) \mid \tag{5}$$

Pokažemo da je $D_n(s)$ res slučajna spremenljivka: Namesto $\sup_{x\in\mathbb{R}}$ lahko zapišemo $\sup_{x\in\mathbb{Q}}$ ker za vsak $x \in \mathbb{R}$ $\exists$ zaporedje $r_i \in \mathbb{Q}$, $i \in \mathbb{N}$, ki konvergira proti x. Torej imamo supremum po števni množici $\Rightarrow D_n$ je slučajna spremenljivka.
Za $1 \leq k \leq n-1$ obstaja, $x_{n,k}$ tako da velja: $F(x_{n,k}) = \frac{k}{n}$. Kadar F ni strogo naraščajoča, je lahko takih števil več, zato je bolje definirati:

$$x_{n,k} = \max\{F^{-1}(\frac{k}{n})\}, \tag{6}$$

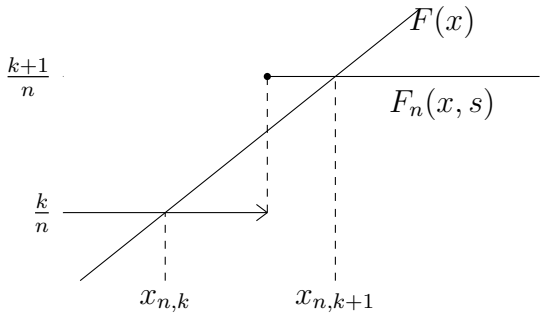
in potem za $x \in (x_{n,k}, x_{n,k+1}]$ velja:

$$\frac{k}{n} \leq F(x) \leq \frac{k+1}{n} \tag{7}$$

Iz slike 2 je razvidno, da je supremum vedno tik pred oz. tik za skokom $F_n(x,s)$. Zaradi preprostosti smo tudi predpostavili, da je F(x) zvezna funkcija.

Dalje je

$$F_n(\alpha^-, s) = lim_{x\to\alpha^-} F_n(x,s) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x)}(X_i(s)).$$



SLIKA 2. empirična porazdelitvena funkcija in $F(x)$ na območju $(x_{n,k}, x_{n,k+1})$

.

Nato definiramo zaporedje $D_n^{(M)}$:

$$D_n^{(M)} := \max_{x \in \Xi_M} \left\{ \mid F_n(x,s) - F(x) \mid, \mid F_n^-(x,s) - F(x) \mid \right\}, \tag{8}$$

kjer je $M \in \mathbb{N}$ in $\Xi_M = \{x_{m,k} \mid 1 \le m \le M, 1 \le k \le m-1\}$. Ker je Ξ končna množica, velja da je

$$S_M = \left\{ s \in S \mid \lim_{n\to\infty} D_n^{(M)}(s) = 0 \right\}$$

dogodek z verjetnostjo 1 po izreku 3.1.
Velja tudi:

$$S_1 \supset S_2 \supset S_3 \supset \ldots \text{ ker } \Xi_1 \subset \Xi_2 \subset \Xi_3 \subset \ldots$$

zato je $S' = \bigcap_{M=1}^{\infty}$ dogodek z verjetnostjo 1.

Izberemo $\epsilon > 0$ in naj bo $\frac{1}{m} < \frac{\epsilon}{2}$ ter $s \in S'$. Potem velja $\lim_{n\to\infty} D_n^{(M)}(s) = 0$, zato $\exists n_o \in \mathbb{N}$, pri katerem je $D_n^{(M)}(s) \le \frac{1}{m}$, za $\forall\ n \ge n_o$. Spodaj obravnavamo primere za vse možne x:

(1) če je $2 \le k \le m$ in $x_{m,k-1} \le x \le x_{m,k}$, velja:

$$\frac{k-1}{m} = F(x_{m,k-1}) \le F(x) \le F(x_{m,k}) = \frac{k}{m}$$

   (a) velja: $F_n(x,s) - F(x) \le F_n(x_{m,k},s) - F(x_{m,k-1}) = F_n(x_{m,k},s) - F(x_{m,k}) + \frac{1}{m}$
   (b) $F(x) - F_n(x,s) \le F(x_{m,k}) - F_n^-(x_{m,k-1},s) = \frac{1}{m} + F_n(x_{m,k}) - F_n(x_{m,k-1},s)$ torej v obeh primerih (a),(b) pod točko (1)$\mid F_n(x,s) - F(x) \mid \le D_n^{(m)}(s) + \frac{1}{m} \le \frac{2}{m} < \epsilon$

(2) za $x \ge x_{m,k}$ velja:
   (a) $F_n(x,s) - F(x) \le 1 - F(x_{m,m-1}) = \frac{1}{m}$
   (b) $F(x) - F_n(x,s) \le 1 - F_n^-(x_{m,m-1},s) = \frac{1}{m} + F(x_{m,m-1}) - F_n(x_{m,m-1},s)$ torej v obeh primerih (a),(b) pod (2)$\mid F_n(x,s) - F(x) \mid \le D_n^{(m)} + \frac{1}{m} < \epsilon$

(3) za $x \le x_{m,1}$ velja:
   (a) $F_n(x,s) - F(x) \le F_n(x_{m,1},s) - 0 = F_n(x_{m,1},s) - F(x_{m,1}) + \frac{1}{m}$
   (b) $F(x) - F_n(x,s) \le F(x_{m,1},s) - 0 = F_n(x_{m,1},s) - F(x_{m,1}) + \frac{1}{m}$ torej v obeh primerih (a),(b) pod točko (3)$\mid F_n(x,s) - F(x) \mid \le D_n^{(m)} + \frac{1}{m} < \epsilon$

$\Rightarrow D_n(s) < \epsilon$ □

4. Vzorec kot slučajni proces

V tem razdelku bom pokazal, kako je možno slučajni vzorec prepoznati kot Poissonov proces, kar bo kasneje ključno pri izračunu verjetnosti $P(D_n \leq d)$. Ideja vzorca kot slučajni proces je bila podana v [1] na strani 2, celotno izpeljavo pa sem izpeljal s pomočjo zapiskov pri predmetu Slučajni procesi 1.

Najprej naredimo verjetnostno integralsko transformacijo na urejenem vzorcu $\vec{S} = (X_{(1)} = x_1, X_{(2)} = x_2, \dots X_{(n)} = x_n)$:

$$t_j = F_0(x_j); j = 1, ..., n,$$

kjer predpostavimo, da je $F_0(x_j)$ zvezna. Ko H_0 drži, je $0 \leq t_1 \leq \cdots \leq t_n$ urejen vzorec n enako porazdeljenih neodvisnih slučajnih spremenjivk, porojenih po $U(0,1)$.

Izrek 4.1. (Verjetnostna integralska transformacija) *Naj bo X slučajna spremenljivka z zvezno porazdelitveno funkcijo $F_X(x)$, potem je slučajna spremenljivka $Y = F_X(X)$ porazdeljena enakomerno na intervalu $(0,1)$.*

Dokaz: če je $0 < y < 1$, potem vzamemo največji x, za katerega je $Y = F_X(x) = y$, obstoj le tega je zagotovljen z zveznostjo porazdelitvene funkcije $F_X(x)$. Tedaj velja $F_X(X) \leq y$ natanko tedaj, ko $X \leq x$. zato $P(Y \leq y) = P(X \leq x) = F_X(x) = y$ $\square$

Primer:

$$H_0: \text{ za vsak } x \in \mathbb{R} F(x) = F_0(x) = exp(1),$$

$$(X_1 = 0.147, X_2 = 0.161, X_3 = 0.329, X_4 = 0.337, X_5 = 0.483,$$
$$X_6 = 0.578, X_7 = 1.066, X_8 = 1.368, X_9 = 2.286, X_{10} = 2.321)$$

$$\tilde{H_0}: \text{ za vsak } x \in \mathbb{R} F(x) = F_0(x) = U(0,1),$$

$$(Y_1 = 0.136, Y_2 = 0.148, Y_3 = 0.280, Y_4 = 0.286, Y_5 = 0.383,$$
$$Y_6 = 0.439, Y_7 = 0.655, Y_8 = 0.745, Y_9 = 0.898, Y_{10} = 0.901)$$

Naj bo $S_t = \sum_{i=1}^{n} \mathbb{1}_{Y_i \leq t}$ funkcija, ki šteje število opazovanj $t_1, \dots, t_n \leq t$ na $0 \leq t \leq 1$. Pokazati želimo, da lahko njeno porazdelitev prepoznamo kot porazdelitev trajektorije Poissonovega procesa, pri katerem smo na intervalu $[0,1]$ opazili n skokov.

Definicija 4.2. *Proces štetja $\{N_t \in \mathbb{N}_0 \mid t \geq 0\}$ je **Poissonov proces** z intenziteto $\lambda > 0$ in skoki velikosti 1, če so izpolnjeni naslednji pogoji :*

(1) $N_0 = 0$ in za $s < t$ velja $N_s \leq N_t$
(2) proces $(N_t)_{t \geq 0}$ ima neodvisne in stacionarne prirastke:
   (a) proces ima neodvisne prirastke, če velja za $\forall t_i, i \in \mathbb{N}_0$,$0 < t_0 < t_1 < \cdots < t_k$, da je slučajni vekor $(N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_k - 1})$ vektor neodvisnih slučajnih spremenjlivk.
   (b) če je izpolnjen še pogoj $(N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_k - 1}) \sim (N_{t_1 + h} - N_{t_0 + h}, N_{t_2 + h} - N_{t_1 + h}, \dots, N_{t_k + h} - N_{t_k - 1 + h})$ za vsak $h \geq 0$ pravimo, da ima proces stacionarne prirastke.
(3) za $\forall t > 0$ je $N_t \sim Pois(\lambda t)$, $P[N_t = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, k = 0, 1, \dots$
(4) za $\forall t$ in $\forall s$, $s < t$ velja $N_t - N_s \sim Pois(\lambda(t - s))$

Lema 4.3. *Naj bo X_j porazdeljen enakomerno na $(0,1)$. Potem se $S_t = \sum_{i=1}^{n} \mathbb{1}_{X_j \leq t}$ $0 \leq t \leq 1$, porazdeljuje enako kot trajektorija prirastkov homogenega Poissonovega procesa, pri pogoju n skokov na intervalu $[0,1]$.*

Dokaz: Namesto trajektorij je dovolj gledati slučajne vektorje $(N_{t_1}, N_{t_2}, \dots N_{t_m}, N_1)$, oziroma vektorje razlik $(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_1 - N_{t_m})$ za vse možne nabore $t_1 \leq t_2 \leq \dots \leq t_m \leq 1$, kjer N_t označuje število skokov do časa t v homogenem Poissonovem procesu.
Oglejmo si pogojno verjetnost:

$$P(N_{t_1} = k_0, N_{t_2} - N_{t_1} = k_1, \dots, N_{t_m} - N_{t_{m-1}} = k_{m-1}, N_1 - N_{t_m} = k_m \mid N_1 = n)$$

Iz pogoja $N_t = n$ vemo, da je $k_0 + k_1 + k_2 + ... + k_m = n$, torej je zgornja pogojna verjetnost enaka:

$$\frac{1}{P(N_1 = n)} P(\bigcap_{i=0}^{m} [N_{t_{i+1}} - N_{t_i}] = k_i) = \frac{n!}{\lambda^n e^{-\lambda}} \prod_{i=0}^{m} \frac{(\lambda(t_{i+1} - t_i))^{k_i} e^{-(\lambda(t_{i+1} - t_i))}}{k_i!} =$$

$$= \frac{n!}{\lambda^n e^{-\lambda}} \lambda^{\sum_{i=1}^{m} k_i} e^{-\lambda \sum_{i=1}^{m} (t_{i+1} - t_i)} \prod_{i=0}^{m} \frac{(t_{i+1} - t_i)^{k_i}}{k_i!} =$$

$$= \frac{n!}{\lambda^n e^{-\lambda}} \lambda^n e^{-\lambda} \prod_{i=0}^{m} \frac{(t_{i+1} - t_i)^{k_i}}{k_i!} =$$

$$= \frac{n!}{k_0! k_1! ... k_m!} \prod_{i=0}^{m} (t_{i+1} - t_i)^{k_i}$$

Pri tem je prva enakost posledica neodvisnosti prirastkov. Ves čas se moramo držati pogoja $k_0 + k_1 + k_2 + ... + k_m = n$, saj je sicer verjetnost enaka 0.
Po drugi strani pa si poglejmo verjetnost $P(X_j \in (t_i, t_{i+1}))$, pri čemer je $X_j \sim U(0,1)$. Potem velja

$$P[X_j \in (t_i, t_{i+1})] = (t_{i+1} - t_i)$$

Od tod in iz neodvisnosti slučajnih spremenljivk X_j dobim :

$$P(S_{t_1} = k_0, S_{t_2} - S_{t_1} = k_2, \dots S_1 - S_{t_m} = k_m) =$$

$$= P[\sum_{i=1}^{n} \mathbb{1}_{X_j \in (0, t_1]} = k_0, \sum_{i=1}^{n} \mathbb{1}_{X_j \in (t_1, t_2]} = k_1, \dots, \sum_{i=1}^{n} \mathbb{1}_{X_j \in (t_{m-1}, t_m)} = k_m] =$$

$$= \frac{n!}{k_0! k_1! ... k_m!} \prod_{1=0}^{m} (t_{i+1} - t_i)^{k_i}$$

Torej smo dokazali, da za vsak končen nabor medčasovnih točk $0 < t_1 < t_2 < \dots < t_m, 1$ in končen n velja, da se $(S_{t_1}, S_{t_2}, \dots, S_{t_m} S_1)$ porazdeljuje enako kot $(N_{t_1}, N_{t_2}, \dots N_{t_m}, N_1)$, ob pogoju n-skokov do časa 1. To pa pomeni, da enako velja tudi za trajektorije. □

5. Izračun dvostranske statistike Kolmogorova in Smirnova

V (4.3) smo pokazali, da se $S_t = \sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$ porazdeljuje enako kot proces štetja pri homogenem poisonovem procesu z intenziteto λ, pri pogoju n skokov na intervalu $[0, 1]$. Potem se empirična porazdelitvena funkcija:

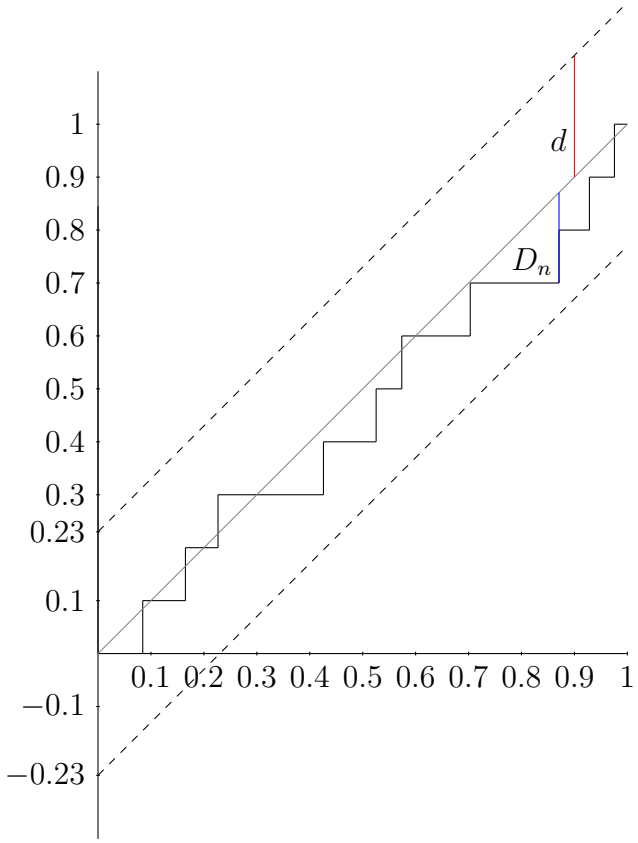
$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$$

porazdeljuje enako kot proces štetja homogenega Poissonovega procesa z intenziteto λ in skoki velikosti $\frac{1}{n}$ pri pogoju n skokov na $[0, 1]$. Poissonov proces s skoki velikosti $\frac{1}{n}$ je definiran kot:

$$P_t = \frac{1}{n} N_t, \text{ kjer je } N_t \in [0, 1] \text{ Poissonov proces z inteziteto } n. \tag{9}$$

Z verjetnostno integralsko transformacijo lahko dosežemo, da je naš vzorec porojen po $U(0, 1)$ in potem velja:

$$\begin{aligned} P(D_n \leq d) &= P(D_n \leq d) = \\ &= P(F_n(t) \text{ leži med premicama} : y = \pm d + t) = \\ &= P(P_t \text{ leži med premicama} : y = \pm d + t \mid P_1 = 1) = \\ &= \frac{P(P_t \text{ leži med premicama} : y = \pm d + t \wedge P_1) = 1)}{P(P_1 = 1)} \end{aligned}$$



Izračun $P(P_t \text{ leži med premicama} : y = \pm d + t \mid P_1 = 1)$ je težaven. Skica izpeljave (brez vmesnih izračunov) je bila podana v [1], na straneh 10-11. Pri izpelavi si z $A_{(j,i)}$ označim vsa možna mesta, kjer lahko trajektorija Poissonovega procesa P_t seka premico $x = \frac{j}{n}$; $j \in (1, 2, ..., n)$ in hkrati ostane med premicama $y = \pm d + t$. Območje

med premicama označim z R. Če nam uspe trajektorijo Poissonovega procesa P_t popeljati iz točke $A(j,i) \to A(j+1,i)$ za vsak j in hkrati ostati v območju R, nam to zagotovi, da D_n ne preseže vrednosti d. Sosledno sem obravnaval ugodne dogodke Poissonovega procesa na intervalu $(\frac{j}{n}, \frac{j+1}{n})$ za prehod iz $A_{(j,i)} \to A_{(j+1,k)}$, pri katerem empirična porazdelitvena funkcija ostane v R. Prirastki homogenega Poissonovega procesa z intenziteto $\lambda > 0$ so neodvisni in stacionarni, obenem velja homogena lastnost Markova, intervali $(\frac{j}{n}, \frac{j+1}{n})_j$ pa so paroma disjunktni, zato lahko izrazim prehode trajektorije P_t iz $(\frac{j}{n}, \frac{j+1}{n}) \to (\frac{j+1}{n}, \frac{j+2}{n})$ s prehodno matriko $H(16)$.

Izrek 5.1. (Homogena lastnost markova): *Naj bo $(N_t)_{t\geq 0}$ homogen Poissonov proces z intenzivnostjo $\lambda > 0$, $f : \mathbb{N}_0 \to \mathbb{R}$ omejena in $0 = t_0 < t_1 < ... < t_l = t$. Potem velja : $E[f(N_{t_n+s}) \mid N_{t_o}, N_{t_1}, ..., N_{t_n}] = E[f(N_{t+s}) \mid N_t]$*

Dokaz:

$$E[f(N_{t_n+s}) \mid N_{t_o}, N_{t_1}, ..., N_{t_n}] \quad = \quad E[f(N_{t_n+s} - N_{t_n} + N_{t_n}) \mid N_{t_o}, N_{t_1}, ..., N_{t_n}]$$

Po prepodsavki homegenga poissonovega procesa so prirastki medseboj neodvisni,zato je tudi $N_{t+s} - N_{t_n}$ neodvisna od $N_{t_o}, N_{t_1}, ..., N_{t_n}$ in dobimo:

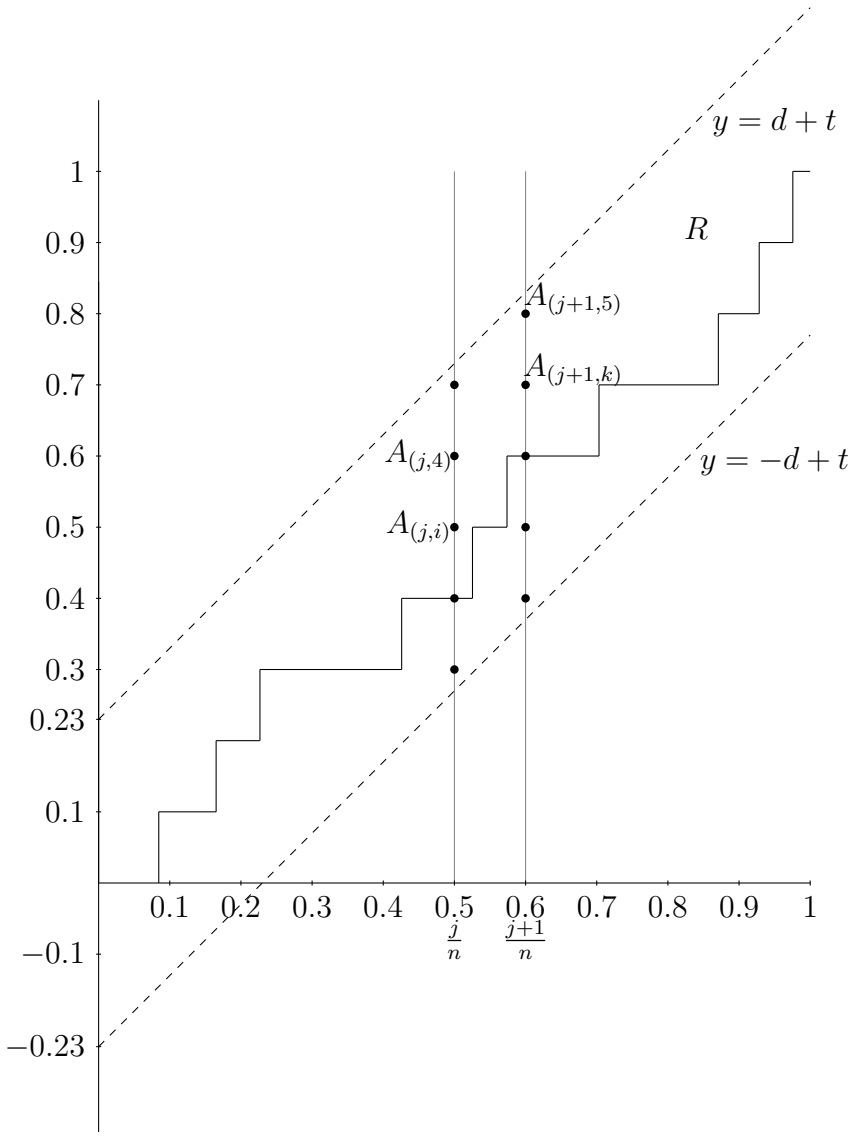
$$\sum_{k=0}^{\infty} f(k + N_{t_n}) e^{-\lambda s} \frac{(\lambda s)^k}{k!}$$

□

V nadaljevanju so prikazane podrobne obravnave verjetnosti da trajektorija P_t ostane v območju R na intervalu $(\frac{j}{n}, \frac{j+1}{n})$. Nadaljna notacija pri izpeljavi:

- r $= \lfloor nd \rfloor$
- $\delta = 1 + r - nd$
- $p = 2r + 1$

5.1. **Verjetnost prehoda** $A_{(j,i)} \to A_{(j+1,k)}$**.** Za prehod iz $A_{(j,i)} \to A_{(j+1,k)}$ mora imeti trajektorija P_t $k-i+1$ skokov na intervalu $(\frac{j}{n}, \frac{j+1}{n})$. V primeru ko je $i,k \in (2,3,\dots p-1)$, in velja $k \geq i-1$.



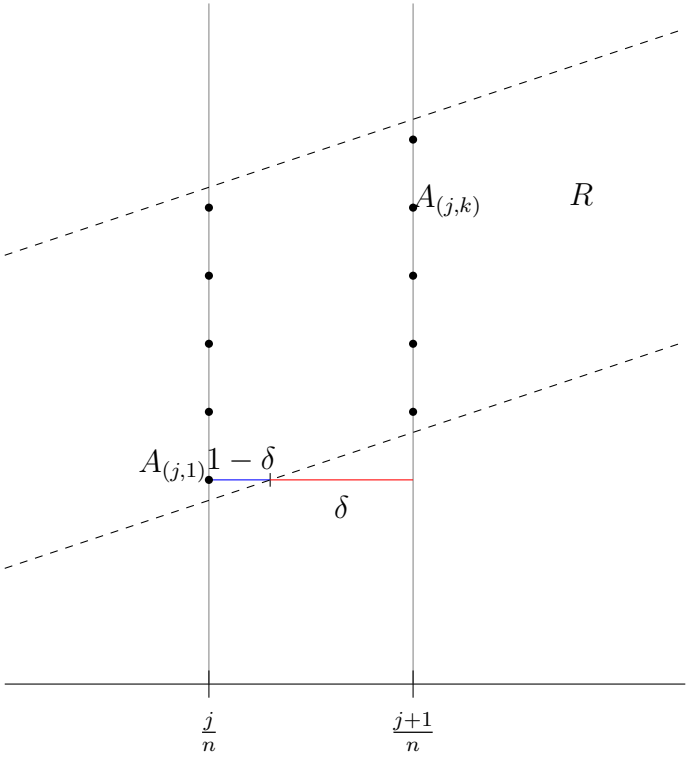
$$t = \frac{j+1}{n} - \frac{j}{n} = \frac{1}{n}$$

$$\begin{aligned} P\left([N_{\frac{j+1}{n}} - N_{\frac{j}{n}}] = k-i+1\right) &= P\left([N_{\frac{1}{n}}] = k-i+1\right) = \\ &= P(Pois(nt) = k-i+1) = \\ &= \frac{(\frac{1}{n}n)^{k-i+1}e^{-nt}}{(k-i+1)!} = \frac{e^{-1}}{(k-i+1)!} \end{aligned}$$

Ker je inteziteta našega procesa $\lambda = n$, opazujemo pa ga vedno na intervalu velikosti $\frac{1}{n}$ oz. njegovi podmnožici, od tu naprej privzamem, da sta:

$$[\lambda = n] \cdot [t = \frac{1}{n}] = 1 \tag{10}$$

5.2. **Verjetnost prehoda** $A_{(j,1)} \to A_{(j+1,k)}$**.** Iz spodnje slike je razvidno, da je iskan dogodek k skokov na intervalu $(\frac{j}{n}, \frac{j+1}{n})$ in hkrati vsaj 1 skok na $(\frac{j}{n}, \frac{j+1-\delta}{n})$, kjer je $k \in (1, 2, \ldots, p-1)$ in $k \geq i-1$, sicer bi naša trajektorija P_t presekala premico $y = -d + t$ in zašla iz območja R.



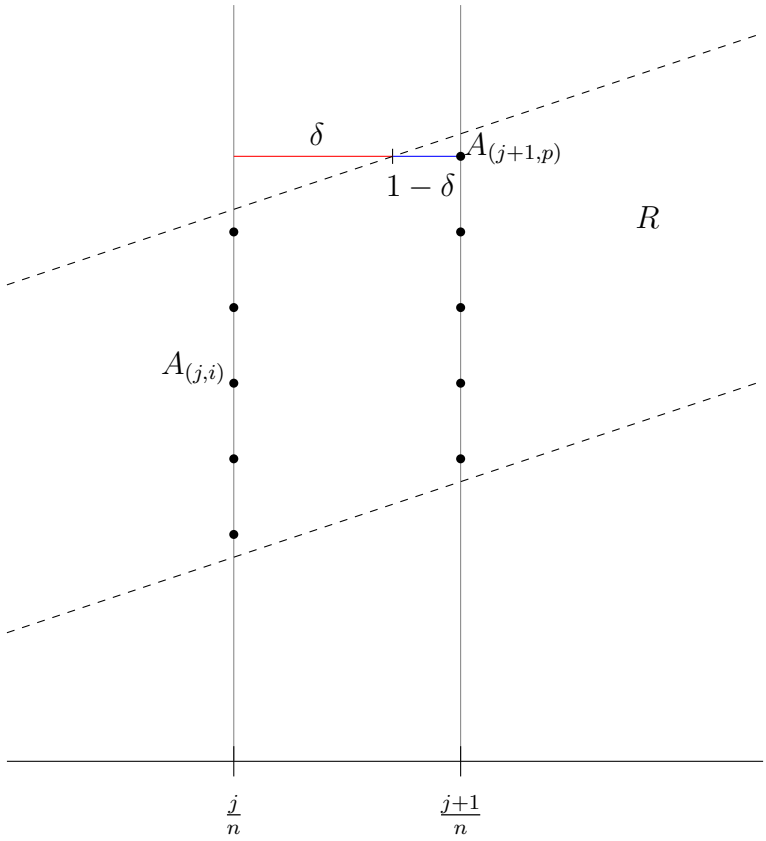
Iskana verjetnost je:

$$\sum_{i=1}^{k} \left[P(N_{\frac{(j+1-\delta)}{n}} - N_{\frac{j}{n}} = i) P(N_{\frac{j+1}{n}} - N_{\frac{j+1-\delta}{n}} = k-i) \right] = \sum_{i=1}^{k} \left[P(N_{\frac{(1-\delta)}{n}} = i) P(N_{\frac{\delta}{n}} = k-i) \right]$$

Upoštevamo (10) in dobimo

$$\begin{aligned}
& \sum_{i=1}^{k} [P(Pois(1-\delta) = i) P(Pois(\delta) = k-i)] = \\
= & \sum_{i=1}^{k} \frac{(1-\delta)^i e^{-(1-\delta)} (\delta)^{k-i} e^{-\delta}}{i!(k-i)!} = e^{-1} \sum_{i=1}^{k} \frac{(1-\delta)^i (\delta)^{k-i}}{i!(k-i)!} = \\
= & e^{-1} \sum_{i=1}^{k} \frac{k!(1-\delta)^i (\delta)^{k-i}}{i!(k-i)!k!} = \frac{e^{-1}}{k!} \sum_{i=1}^{k} \binom{k}{i} (1-\delta)^i (\delta)^{k-i} = \\
= & \frac{e^{-1}}{k!} \left[\sum_{i=0}^{k} \binom{k}{i} (1-\delta)^i (\delta)^{k-i} - \frac{k!(1-\delta)^0 \delta^k}{k!0!} \right] = \\
= & \frac{e^{-1}}{k!} [(1-\delta+\delta)^k - \delta^k] = \frac{e^{-1}(1-\delta^k)}{k!}
\end{aligned}$$

5.3. **Verjetnost prehoda** $A_{(j,i)} \to A_{(j+1,p)}$**.** Imamo $p-i+1$ skokov na intervalu $(\frac{j}{n}, \frac{j+1}{n})$, kjer je $i > 1$. Pri tem mora biti vsaj en skok $\in (\frac{j+\delta}{n}, \frac{j+1}{n})$, saj v primeru, ko so vsi skoki $\in (\frac{j}{n}, \frac{j+\delta}{n})$ P_t, preseka zgornjo premico $y = -d + t$, in tako zaide iz območja R.



$$\sum_{j=1}^{p-i+1} \left[P(N_{\frac{(j+\delta)}{n}} - N_{\frac{j}{n}} = p-i+1-j) P(N_{\frac{j+1}{n}} - N_{\frac{j+\delta}{n}} = j) \right] =$$

$$= \sum_{j=1}^{p-i+1} \left[P(N_{\frac{(\delta)}{n}} = i) P(N_{\frac{1-\delta}{n}} = j) \right] =$$

$$= \sum_{j=1}^{p-i+1} \left[P(Pois(\delta) = \{p-i+1-j\}) P(Pois(1-\delta) = \{j\}) \right] =$$

$$= \sum_{j=1}^{m=p-i+1} \frac{(\delta)^{m-j} e^{-\delta} (1-\delta)^j e^{-(1-\delta)}}{(m-j)!(j)!} = e^{-1} \sum_{j=1}^{m=p-i+1} \frac{m!(\delta)^{m-j}(1-\delta)^j}{(m-j)!(j)!m!} =$$

$$= \frac{e^{-1}}{m!} \sum_{j=1}^{m=p-i+1} \binom{m}{j} (\delta)^{m-j} (1-\delta)^j =$$

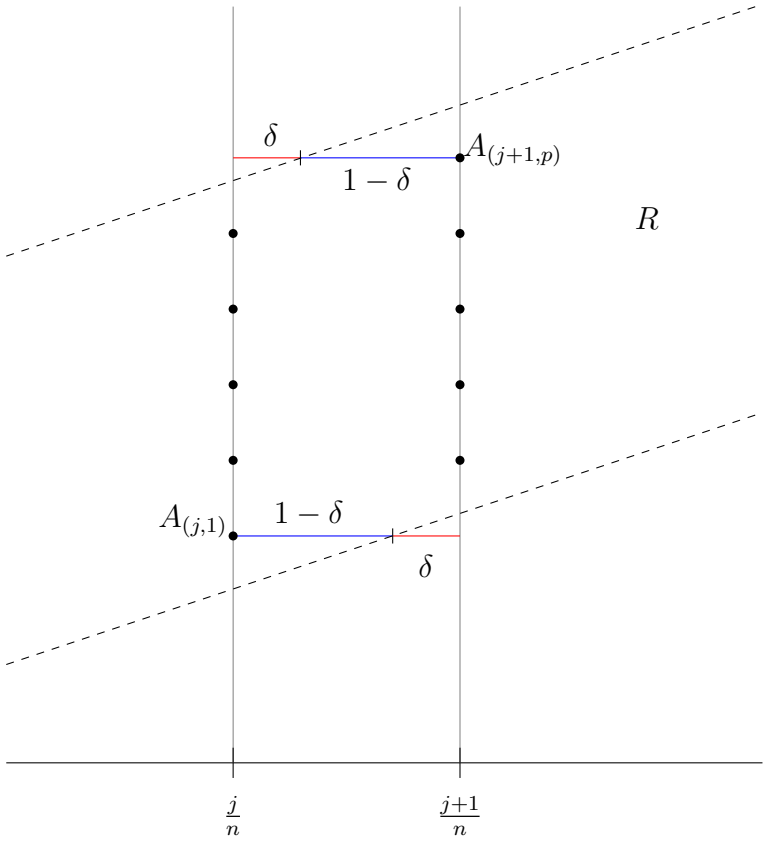
$$= \frac{e^{-1}}{m!} \left[\sum_{i=0}^{m} \binom{m}{j} (\delta)^{m-j} (1-\delta)^j - \frac{m!(\delta)^m (1-\delta)^0}{m!0!} \right] =$$

$$= \frac{e^{-1}}{m!} \left[(\delta + 1 - \delta)^m - \delta^m \right] = \frac{e^{-1}(1-\delta^{p-i+1})}{(p-i+1)!}$$

5.4. **Verjetnost prehoda** $A_{(j,1)} \to A_{(j+1,p)}$**.** Tukaj obravnavam še primer, ko P_t skoči iz $A_{(j,1)}$ na sam vrh dovoljenega območja $A_{(j+1,p)}$. Torej se zgodi p skokov na

$(\frac{j}{n}, \frac{j+1}{n})$. Za lajšanje izračuna sem problem razčlenil na dva podprimera: $\delta \leq \frac{1}{2}$ in $\delta > \frac{1}{2}$.

(1) $\delta \leq \frac{1}{2}$: Iščem dogodek s p skoki na intervalu $(\frac{j}{n}, \frac{j+1}{n})$, pri tem je vsaj en skok $\in (\frac{j}{n}, \frac{j+1-\delta}{n})$ in hkrati niso vsi $\in (\frac{j}{n}, \frac{j+\delta}{n})$, saj bi to pomenilo, da smo že zašli iz območja R.



Iskano verjetnost izrazimo kot p skokov na $(\frac{j}{n}, \frac{j+1}{n})$ pri tem vsaj en $\in (\frac{j}{n}, \frac{j+1-\delta}{n})$, kar ustreza:

$$\sum_{i=1}^{p} P\left[(N_{\frac{(j+1-\delta)}{n}} - N_{\frac{j}{n}} = i)(N_{\frac{j+1}{n}} - N_{\frac{(j+1-\delta)}{n}} = j)\right] = \frac{e^{-1}(1-\delta^p)}{p!} \tag{11}$$

Zgornjo vsoto izračunamo na enak način kot v podpoglavju 5.3 , ter naknadno odštejemo verjetnost, da so bili vsi skoki $\in (\frac{j}{n}, \frac{j+\delta}{n})$.

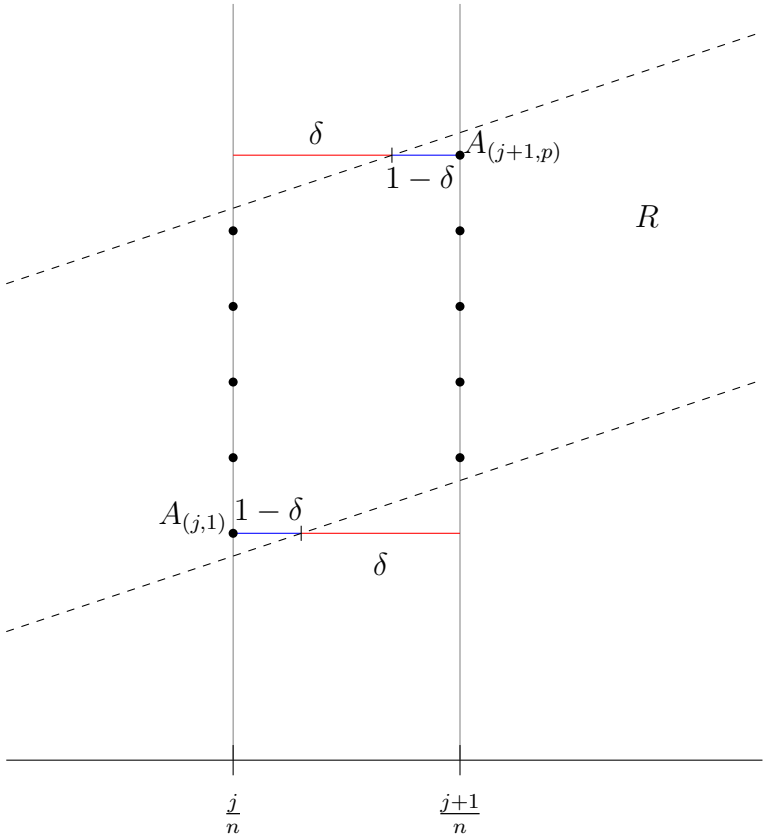
$$P\left[(N_{\frac{(j+\delta)}{n}} - N_{\frac{j}{n}} = p)(N_{\frac{(j+1)}{n}} - N_{\frac{(j+1-\delta)}{n}} = 0)\right] = P\left[(N_{\frac{(\delta)}{n}} = p)(N_{\frac{(1-\delta)}{n}} = 0)\right] \tag{12}$$

Odštejemo izraza (11) in (12), ter dobimo

$$\frac{e^{-1}(1-\delta^p)}{p!} - P(Pois(\delta) = [p])(Pois(1-\delta) = [0]) = = \frac{e^{-1}(1-\delta^p)}{p!} - \frac{e^{-1}(\delta^p)}{p!} = = \frac{e^{-1}(1-2\delta^p)}{p!}$$

(2) $\delta > \frac{1}{2}$: naš iskan dogodek je enak kot v (1), vendar je tokrat območje $[\frac{j}{n}, \frac{j+1-\delta}{n}] \subset [\frac{j}{n}, \frac{j+\delta}{n}]$. Če ponovno uporabimo zgornjo formulo, ne vemo točno koliko skokov se je pripetilo na intervalu $\in (\frac{j+1-\delta}{n}, \frac{j+\delta}{n})$ in posledično štejemo

še načine dogodka, ko je vzorčna pot že zašla iz območja R. V izogib dvojnim vsotam lahko te dogodke odštejem:



Iskano verjetnost izrazimo kot p skokov na intervalu $(\frac{j}{n}, \frac{j+1}{n})$, pri tem vsaj en $\in (\frac{j}{n}, \frac{j+1-\delta}{n})$ in dobimo enako kot pri izrazu (11):

$$\sum_{i=1}^{p} P\left[(N_{\frac{(j+1-\delta)}{n}} - N_{\frac{j}{n}} = i)(N_{\frac{j+1}{n}} - N_{\frac{(j+1-\delta)}{n}} = j)\right] = \frac{e^{-1}(1-\delta^p)}{p!} \tag{13}$$

naknadno odštejem še način zgornjega dogodka, pri katerem so vsi skoki $\in [\frac{j}{n}, \frac{j+\delta}{n}]$.

$$\sum_{i=1}^{p} P\left[(N_{\frac{(j+1-\delta)}{n}} - N_{\frac{j}{n}} = i))(N_{\frac{(j+\delta)}{n}} - N_{\frac{(j+1-\delta)}{n}} = p-i)](N_{\frac{(j+1)}{n}} - N_{\frac{(j+\delta)}{n}} = 0)\right] \tag{14}$$

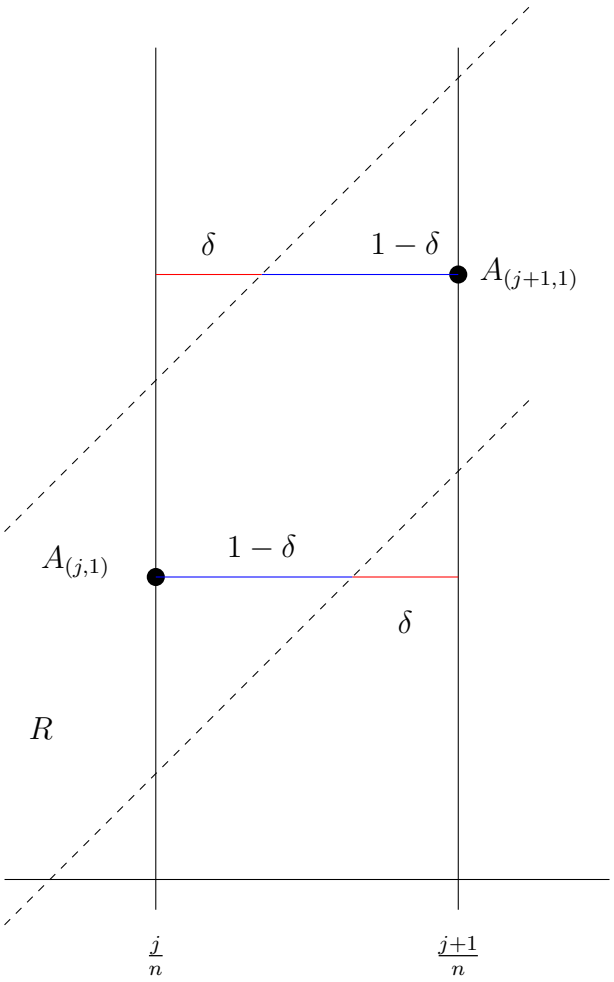
od (13) odštejemo (14) in dobimo

$$
\begin{aligned}
&= \frac{e^{-1}(1-\delta^p)}{(p)!} - \sum_{i=1}^{p} \frac{(1-\delta)^i e^{-(1-\delta)}(2\delta-1)e^{-(2\delta-1)}(1-\delta)^0 e^{-(1-\delta)}}{i!(p-i)!} = \\
&= \frac{e^{-1}(1-\delta^p)}{(p)!} - e^{-1}\sum_{i=1}^{p} \frac{(1-\delta)^i(2\delta-1)^{p-i}}{i!(p-i)!} = \\
&= e^{-1}\left[\frac{(1-\delta^p)}{(p)!} - \sum_{i=1}^{p} \frac{p!(1-\delta)^i(2\delta-1)^{p-i}}{i!(p-i)!p!}\right] = \\
&= \frac{e^{-1}}{p!}\left[1-\delta^p - \sum_{i=1}^{p}\binom{m}{j}(1-\delta)^i(2\delta-1)^{p-i}\right] = \\
&= \frac{e^{-1}}{p!}\left[1-\delta^p - \left(\sum_{i=0}^{p}\binom{m}{j}(1-\delta)^i(2\delta-1)^{p-i} - (2\delta-1)^p\right)\right] = \\
&= \frac{e^{-1}}{p!}\left[1-\delta^p - (\delta^p - (2\delta-1)^p)\right] = \\
&= \frac{e^{-1}(1-2\delta^p + (2\delta-1)^p)}{p!}
\end{aligned}
$$

5.5. **Verjetnost prehoda** $A_{(j,1)} \to A_{(j+1,1)}$**.** Naj omenim še robni primer, ko je $nd < 1$. Potem velja

$$r = \lfloor nd \rfloor = 0, p = 2r + 1 = 1.$$

Imamo le eno možno mesto, kjer vzorčna pot lahko seka $\frac{j+1}{n}$ in hkrati ostane znotraj območja R. Imamo 1 skok na $(\frac{j+\delta}{n}, \frac{j+1-\delta}{n})$, 0 skokov na $(\frac{j}{n}, \frac{j+\delta}{n})$ in 0 skokov na $(\frac{j+1-\delta}{n}, \frac{j+1}{n})$, kar ustreza:



Iskana verjetnost je enaka:

$$\begin{aligned}
& P\left([N_{\frac{j+\delta}{n}} - N_{\frac{j}{n}}] = 0\right) P\left([N_{\frac{j+1-\delta}{n}} - N_{\frac{j+\delta}{n}}] = 1\right) P\left([N_{\frac{j+1}{n}} - N_{\frac{j+1-\delta}{n}}] = 0\right) = \\
= \;& P\left([N_{\frac{\delta}{n}}] = 0\right) P\left([N_{\frac{1-2\delta}{n}}] = 1\right) P\left([N_{\frac{\delta}{n}}] = 0\right) = \\
= \;& P(Pois(\delta) = \{0\}) P(Pois(1-2\delta) = \{1\}) P(Pois(\delta) = \{0\}) = \\
= \;& \frac{(\delta)^0 e^{-(\delta)} (1-2\delta)^1 e^{-(1-2\delta)} (\delta)^0 e^{-(\delta)}}{(0)!(1)!(0)!} = e^{-1}(1-2\delta)
\end{aligned}$$

Rezultat je enak kot v 5.4. Iz zgornje slike razberemo, da v primeru, ko je $\delta > \frac{1}{2}$ in $nd < 1$, vzorčna pot nima možnosti za obstoj znotraj območja R, torej je v tem primeru $P(D_n \leq d) = 0$. Kar se zgodi natanko tedaj, ko velja:

$$\begin{aligned}
\delta = 1 + r - nd > \frac{1}{2} \qquad & r = \lfloor nd \rfloor = 0 \\
\frac{1}{2} > nd \qquad &
\end{aligned}$$

Rezultat je intiutivno smiseln, saj D_n ne mora biti poljubno majhen pri končnem n.

Za lažjo obravnavo in izračun vse skupaj ponazorim v prehodni matriki $H \in \mathbb{R}^{p\times p}$. Za $p > 1$ velja:

$$H = \begin{bmatrix} 1-\delta & 1 & 0 & 0 & \cdots & 0 \\ \frac{1-\delta^2}{2!} & 1 & 1 & 0 & \cdots & 0 \\ \frac{1-\delta^3}{3!} & \frac{1}{2!} & 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1 \\ \frac{1-2\delta^p+(2\delta-1)_+^p}{p!} & \frac{1-\delta^{p-1}}{(p-1)!} & \frac{1-\delta^{p-2}}{(p-2)!} & \cdots & \frac{1-\delta^2}{2!} & 1-\delta \end{bmatrix} \tag{15}$$

- $H_{1,1}, H_{1,2}, ..., H_{1,p-1}$ predstavlja verjetnost $A_{(j,1)} \to A_{(j+1,k)}$,
- $H_{1,p} : A_{(j,1)} \to A_{(j+1,p)}$,
- $H_{p,2}, H_{p,3}, ..., H_{p,p} : A_{(j,i)} \to A_{(j+1,p)}$,
- $H_{i,k}, : A_{(j,i)} \to A_{(j+1,k)}$ za $k, i \in (2, 3, ..., p-1)$, hrati pa $k \geq i-1$,
- ostala mesta predstavljajo negativno število skokov, zato je na teh mestih verjetnost enaka 0.

Ko je $p = 1$ in $nd > \frac{1}{2}$, je matrika $H \in \mathbb{R}^{1\times 1}$.

$$H = \begin{bmatrix} 2\delta - 1 \end{bmatrix}$$

Naj bo $e^{-j}u_{ji}$ verjetnost, da P_t doseže točko A_{ji} ter hkrati ostane v dovoljenem območju R, kjer je u_j vektor stanj v koraku j.

$$u_j = [u_{j1}, u_{j2}, ..., u_{jp}], j = 0, 1, ...n, p = 2r - 1, \tag{16}$$

Potem lahko prehod med stanji podamo z relacijo $u_{j+1} = Hu_j$, kjer je u_0 ničelni vektor z vrednostjo 1 v $r + 1$, saj to predstavlja začetek vzočne poti v točki $(0, 0)$.

Primer:

$$\begin{bmatrix} 1-\delta & 1 & 0 & 0 & 0 \\ \frac{1-\delta^2}{2!} & 1 & 1 & 0 & 0 \\ \frac{1-\delta^3}{3!} & \frac{1}{2!} & 1 & 1 & 0 \\ \frac{1-\delta^4}{24} & \frac{1}{6} & \frac{1}{2!} & 1 & 1 \\ \frac{1-2\delta^5+(2\delta-1)_+^5}{125} & \frac{1-\delta^4}{24} & \frac{1-\delta^3}{6} & \frac{1-\delta^2}{2} & 1-\delta \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ \frac{1}{2!} \\ \frac{1-\delta^3}{6} \end{bmatrix} \tag{17}$$

Z večkratnim ponavljanjem (17) dobimo $u_n = H^n u_o$. Verjetnost, da trajektorija našega Poissonovega procesa konča v točki $(1, 1)$ ter je vseskozi v območju R, je torej enaka $e^{-n} \cdot ([r + 1], [r + 1])$-ti element H^n. Brezpogojna verjetnost je enaka :

$$P(N_1 = n) = \frac{e^{-n}(n)^n}{n!} \tag{18}$$

Relacijo $e^{-n} \cdot ([r + 1], [r + 1])$-ti element H^n združimo z brezpogojno verjetnostjo (18) in dobimo :

$$\begin{aligned} P(D_n \leq d) &= P(D_n < d) = \\ &= \frac{P(P_t \text{ leži med premicama} : y = \pm d + t \wedge P_1 = 1)}{P(P_1 = 1)} = \\ &= \frac{n!}{n^n} * ([r + 1], [r + 1])\text{-ti element matrike } H^n. \end{aligned}$$

Primer:

$P(D_{10} \leq 0.394)$

izrazimo $d = \frac{1-\delta+r}{n}$, kjer je r naravno število, $\delta \in (0, 1)$:

(1) $r = 3$
(2) $\delta = 0.06$
(3) $p = 2r + 1 = 7$

V tem primeru je $\delta < \frac{1}{2}$, $p = 7$, torej $H \in \mathbb{R}^{7\times 7}$

$$H = \begin{bmatrix} 1-\delta & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1-\delta^2}{2!} & 1 & 1 & 0 & 0 & 0 & 0 \\ \frac{1-\delta^3}{3!} & \frac{1}{2!} & 1 & 1 & 0 & 0 & 0 \\ \frac{1-\delta^4}{4!} & \frac{1}{3!} & \frac{1}{2!} & 1 & 1 & 0 & 0 \\ \frac{1-\delta^5}{5!} & \frac{1}{4!} & \frac{1}{3!} & \frac{1}{2!} & 1 & 1 & 1 \\ \frac{1-\delta^6}{6!} & \frac{1}{5!} & \frac{1}{4!} & \frac{1}{3!} & \frac{1}{2!} & 1 & 0 \\ \frac{1-2\delta^7}{7!} & \frac{1-\delta^6}{6!} & \frac{1-\delta^5}{5!} & \frac{1-\delta^4}{4!} & \frac{1-\delta^3}{3!} & \frac{1-\delta^2}{2} & 1-\delta \end{bmatrix}$$

$P(D_{10} \leq 0.394) = \frac{10!}{10^{10}} H^{10} = 0.9344877$

Če je prepodstavljen $\alpha >$ P-vrednost, lahko H_0 zavrnemo pri P-vrdnosti:

$$\text{P-vrednost} = 1 - 0.9344877 = 0.0655123$$

6. Algoritem v jeziku R za izračun $P(D_n \leq d)$

Izdelal sem tudi algoritem v programskem jeziku R, ki izračuna $P(D_n <= d)$ na 7 decimalnih mest natančno za n velikosti od 2 do 300. Kar v praksi zadostuje, saj se že za n-je od 200 naprej uporablja limitna oblika

$$\lim_{n\to\infty} P(\sqrt{n}D_n \leq x) = L(x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2} = \frac{\sqrt{2}\pi}{x}\sum_{i=1}^{\infty} e^{-(2i-1)^2\pi^2/(8x^2)},$$

pri čemer je prvo reprezentacijo podal sam Kolmogorov, druga pa prihaja iz standardne realizacije za theta funkcije in je primernejša za majhne x. Obe pa sta bili podani v članku [5].

```
library(Rmpfr)
Potenciranje <- function(A, n) {
#ekonomičen izračun potenc matrike H
  oznake <- c()
  while (n>1) {
    if (n%%2==0) {
      oznake <- c(0, oznake)
      n <- n%/%2
    } else {
      oznake <- c(1, oznake)
      n <- n - 1
    }
  }
  B <- A
  for (i in oznake) {
    if (i) {
      B <- B%*%A
    } else {
      B <- B%*%B
    }
  }
```

```
  return(B)
}
Dn <- function(d, n) {
if (n*d < 0.5) {
  return(0)
}
if (n*d > 60) {
  return(1)
}
  #izracunamo r, delta in p
  x <- d*n
  r <- floor(x)
  delta <- 1 + r - x
  p <- 2*r + 1
  #skonstruiramo H
if (n*d >= 1) {
  H <- matrix(0, nrow=p, ncol=p)
  for (i in 1:(p-1)) {
    for (j in 2:(i+1)) {
      H[i, j] <- 1/factorial(i-j+1)
     #predstavlja verjetnost prehoda A(j,i)->A(j+1,k)
    }
    H[i, 1] <- (1-deltaˆi)/factorial(i)
    #prvi stolpec brez zadnjega elementa predstavlja A(j,1)->A(j+1,k)
    H[p, p+1-i] <- (1-deltaˆi)/factorial(i)
    #zadnja vrstica brez prvega elementa predstavlja A(j,i)->A(j+1,p)
  }
  #p,1-ti element v odvisnosti od delta predstavlja A(j,1)->A(j+1,p)
  if (delta <= .5) {
    H[p, 1] <- (1-2*deltaˆp)/factorial(p)
  } else {
    H[p, 1] <- (1-2*deltaˆp+(2*delta-1)ˆp)/factorial(p)
  }
}else {
H <- (1- 2*delta)
}
  a<-1
   for (i in 1:n) {
  a <- a * (i/n)
  }
  #izračunamo Hˆn
  T <- Potenciranje(H, n)
  t <- T[r+1, r+1] #(r+1, r+1) ti element Hˆn
  return(a*t)
}
Dn(.394, 10)
[1] 0.9344877
>
```

Funkcija potenciranje ima v najboljšem primeru $log_2(n)$ množenj, kar se zgodi, ko je n oblike 2^i, v skrajnem primeru pa $2log_2(n)$, ko je n oblike $(((2+1)^2+1)^2\cdots)^2$.

Primer:

$$n = 100, \text{oznake} = [01000100]H^n = (((((((H^2)H)^2)^2)^2)H)^2)^2.$$

Množenje matrik ima navadno časovno zahtevnost $O(p^3)$. V resnici predstavlja največjo težavo $\frac{n!}{n^n}$. Zadevo skušam omiliti z zanko, ki $\frac{n!}{n^n}$ izrazi kot $\frac{1}{n} * \frac{2}{n} * \ldots \frac{n}{n}$. Vsekakor to ni optimalen pristop k reševanju problema. Kolikor sem informiran je najmočnejši algoritem, ki izračuna $P(D_n \leq n)$ na vsaj 10 decimalnih mest natačno in za n-je do velikosti 16000, podan v članku [5].

Literatura

[1] J. Durbin, *Distribution Theory for tests Based on the Sample Distribution Function*, Society for Industrial and Applied Mathematics, Philadelphia, 1972, strani 1–12.

[2] M. Hladnik, *Verjetnost in Statistika*, Založba fakultete za elektrotehniko in fakultete za računalništvo in informatiko, Ljubljana, 2002.

[3] R. Jamnik, *Verjetnostni račun in Statistika*, Založba fakultete za elektrotehniko in fakultete za računalništvo in informatiko, Ljubljana, 2002.

[4] A. Kolmogorov, *Sulla determinazione empirica di una legge di distributione*, Giornale dell' Istituto Italiano degli Attuari 4, (1933) strani 83–91.

[5] G. Marsaglia, W. W. Tsang, J. Wang *Evaluating Kolmogorov's Distribution*, Journal of Statistical Software 8, (2003), št.15