8. Generatorji naključnih števil

Domača naloga pri predmetu Modelska analiza I

Avtor: Matic Noč

15.11.2017

1 Uvod

Obravnavamo generatorje naključnih števil. Naključna števila bodo pri dobrem generatorju vedno porazdeljena enakomerno. Ker pa potrebujemo naključna števila tudi v različnih porazdelitvah, geometrijskih oblikah, lahko s transformacijami dveh naključnih števil izračunamo nova naključna števila, ki so drugače porazdeljena. Ogledali si bomo različno porazdeljena naključna števila, preverili njihove porazdelitve s statističnimi testi in nato obravnavali statistično analizo oddaje nalog modelske analize.

2 Generator naključnih števil

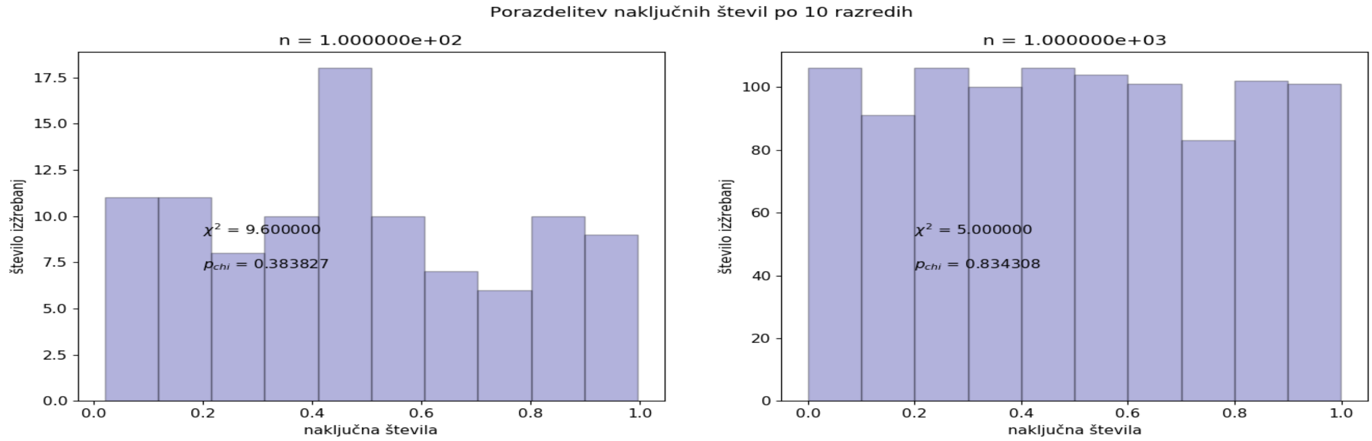
Preverimo najprej delovanje generatorja naključnih števil. Uporabili bomo pseudorandom Mersenne Tvister (1997 by Makoto Matsumoto in Nishimura), ki ima periodo $2^{19937}-1$ in serijsko enakomernost do 623 števil.

2.1 Enakomerna porazdelitev naključnih števil

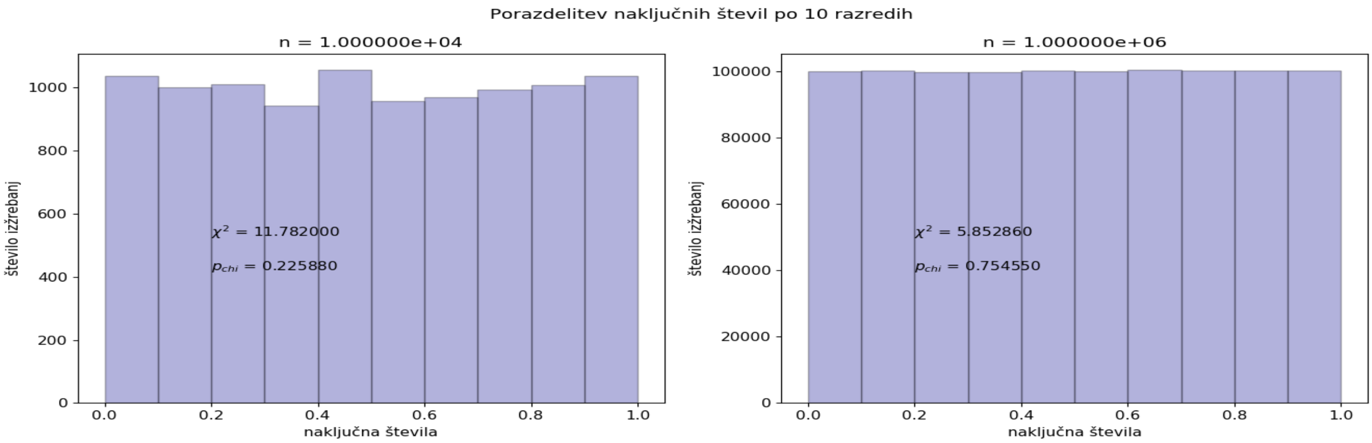
Če računamo naključna števila med 0 in 1 moramo dobiti enakomerno porazdelitev števil. Poglejmo si histograme in statistične teste generatorja Marsenne Twister. Najprej testiramo s chi^2 statistiko:

$$\chi^2 = \sum_i^n \frac{(O_i - E_i)^2}{E_i^2} \tag{1}$$

kjer je i parameter razreda, O izmerjeno/zaznano število spremenljivke v določenem razredu in E pričakovano število spremenljivke v i-tem razredu. V primeru enakomerne porazdelitve je $E_i = \int_{x_i}^{x_{i+1}} \frac{dP}{dx} dx = \mathbf{\frac{1}{n} N}$, kjer je N število vseh izmerkov.

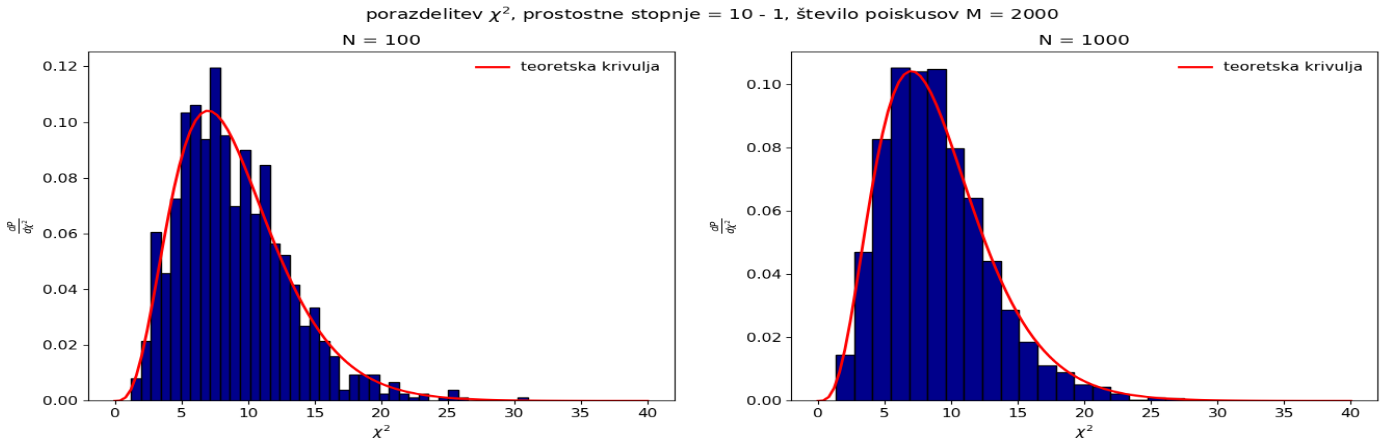


Slika 1: Histogram 10 razredov za različno število naključnih žrebov (n=100,1000). Vidimo, da z višanjem števil vse bolj zapolnimo prostor in imamo res enakomerno porazdelitev.

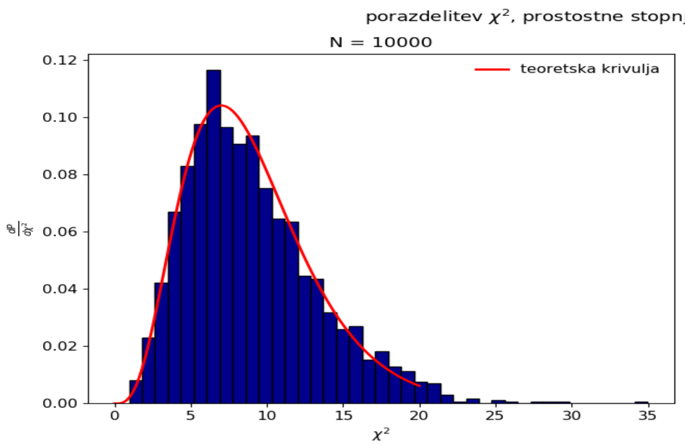


Slika 2: Histogram 10 razredov za različno število naključnih žrebov. Vidimo, da z višanjem števil vse bolj zapolnimo prostor in imamo res enakomerno porazdelitev. Vidimo, da je za vse n zadoščen pogoj $\chi^2 < \chi^2_D$ oziroma $p > 0.05$.

Vsakič ko naključno vlečemo števila seveda χ^2 ne pride isti. Porazdeljen je po analitičnem porazdelitvenem zakonu $\frac{dP}{d\chi^2}$ za $n_{razredov} - 1$ prostostnih stopenj. V primeru da je naša porazdelitev res enaka (ničta hipoteza H0) bodo χ^2 morali nanizati porazdelitev za ustrezne prostostne stopnje.



Slika 3: Po 2000 poiskusih naključnega žreba 100, 1000 števil vidimo da je hi-kvadrat res porazdeljen po ustreznem zakonu s povprečjem $n_{razredov} - 1$, in še en dokaz, da naša hipoteza o enakosti porazdelitev drži.



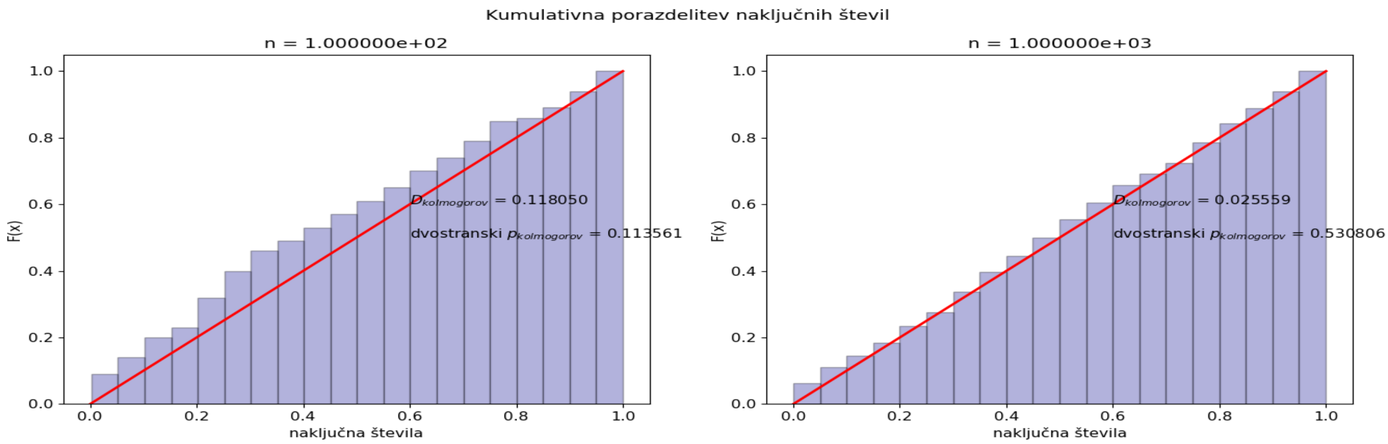
Slika 4

2.2 Test Kolmogorov - Smirnov

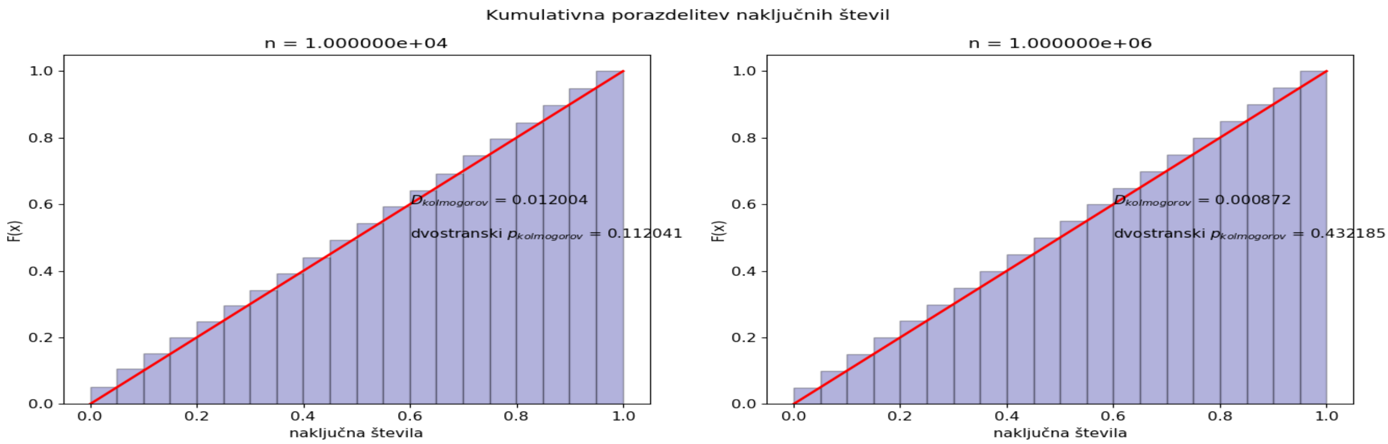
Opravimo lahko tudi test kolmogorova, kjer primerjamo dve kumulativni porazdelitvi, tako da iščemo največjo razdaljo med kumulativnima porazdelitvama

$$D = \sup_x |F_n(x) - F(x)|, \tag{2}$$

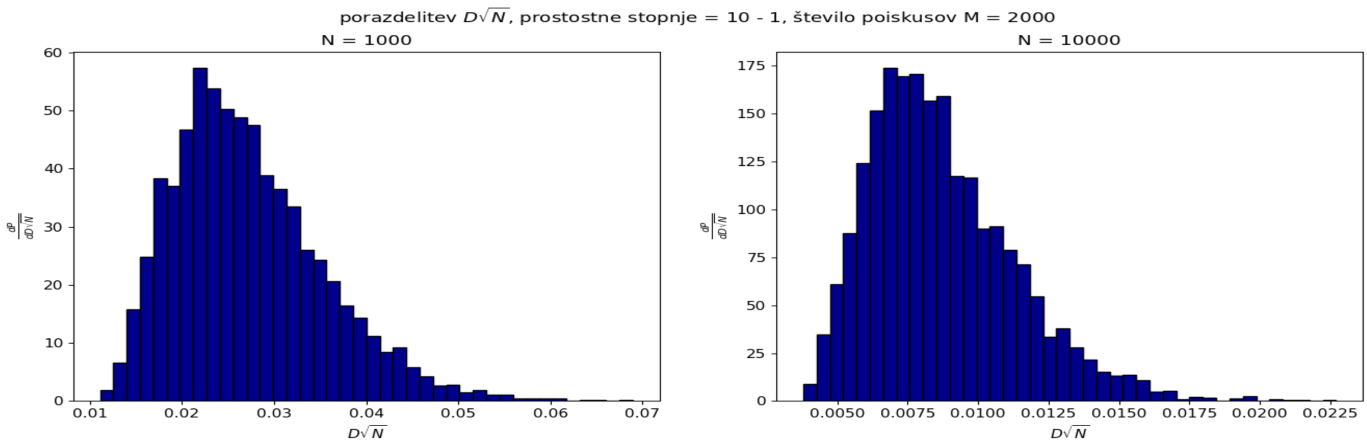
kjer je $F_n(x)$ izmerjena kumulativna porazdelitev in $F(x)$ testna porazdelitev. Kolmogorov je dokazal, da so $D\sqrt{N}$,kjer je (N število izmerkov), porazdeljeni po zakonu Kolmogorova, in zato lahko testiramo enakost porazdelitev. Če sta porazdelitvi res enaki, bodo morali biti za dane prostostne stopnje, $D\sqrt{N}$ porazdeljeni po zakonu Kolmogorova. Če pa smo porazdelitev zgrešili pa bodo D veliko večji in tako ne bodo porazdeljeni po zakonu Kolmogorova za dane prostostne stopnje. To lahko zopet simuliramo s M ponovitvami žrebov naključnih števil in vsakič izračunamo test Kolmogorova med žrebom in enakomerno porazdelitvijo.



Slika 5



Slika 6: Vidimo, da je pri vseh razdalja dovolj velika, da je levo in desno od aboslutne vrendosti razdalje verjetnost $p > 0.05$.



Slika 7: Porazdelitev statistike ustreza porazdelitvi kolmogorova.

3 Normalno porazdeljena naključna števila

Za simulacije večinoma ne potrebujemo enakomerno porazdeljenih naključnih števil, temveč nekakšno drugo porazdelitev npr. normalno.

3.1 Box Muller

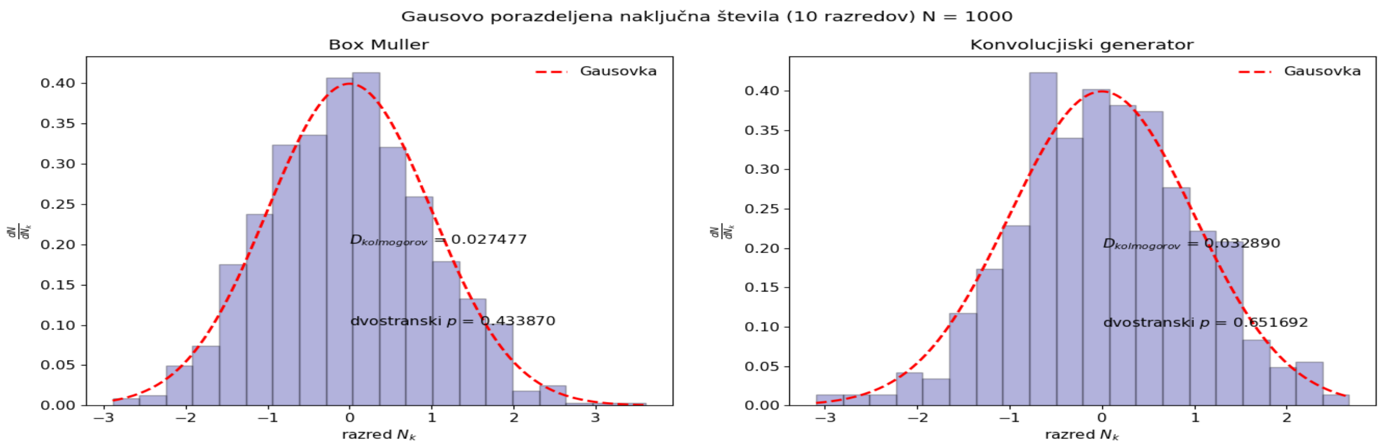
Box - Muller transformacija nam dve enakomerno porazdeljenih naključnih števili transformira v normalno porazdeljeni naključni števili s transformacijo.

$$\begin{aligned} Z_1 &= R\cos(\Theta) = \sqrt{-2\ln U_1}\cos(2\pi U_2), \\ Z_2 &= R\sin(\Theta) = \sqrt{-2\ln U_1}\sin(2\pi U_2). \end{aligned} \tag{3}$$

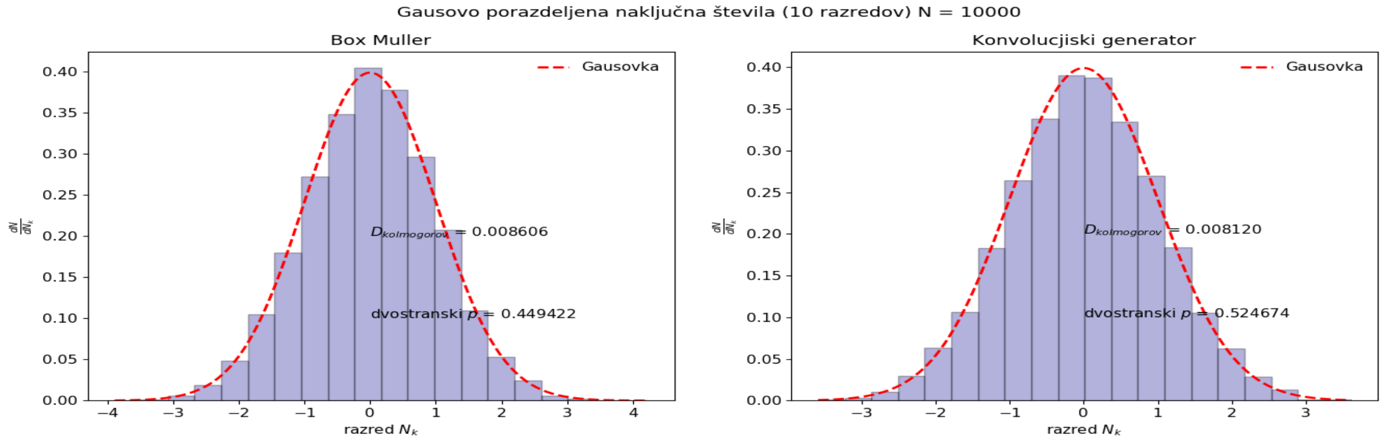
Drug način pa je konvolucijski tip generatorja, kjer vzamemo konvolucijo enakmerne porazdelitve in po centralnem limitnem teoremu preidemo v normalno porazdelitev. Tako vzamemo npr. 6 naključnih števil iz random generatorja in jih seštejemo, ter nato 6 naključnih števil odštejemo, da bo normalna porazdelitev imela povprečje 0 in sigmo ena.

$$Z = \sum_i^6 X_i - \sum_i^6 Y_i \tag{4}$$

Očitno je da je konvolucijski algoritem počasnejši saj potrebuje 6 krat več računati naključna števila kot Box Muller.



Slika 8: Naključno generirana števila (N= 1000) porazdeljena po normalni porazdelitvi. Porazdelitev smo testirali s testom Kolmogorov - Smirnov, in očitno je $p > 0.05$, torej lahko potrdimo normalno porazdelitev.



Slika 9: Naključno generirana števila (N= 10000) porazdeljena po normalni porazdelitvi. Pri višanju žrebov se razredi vse bolj polnijo in je porazdelitev vedno lepša.

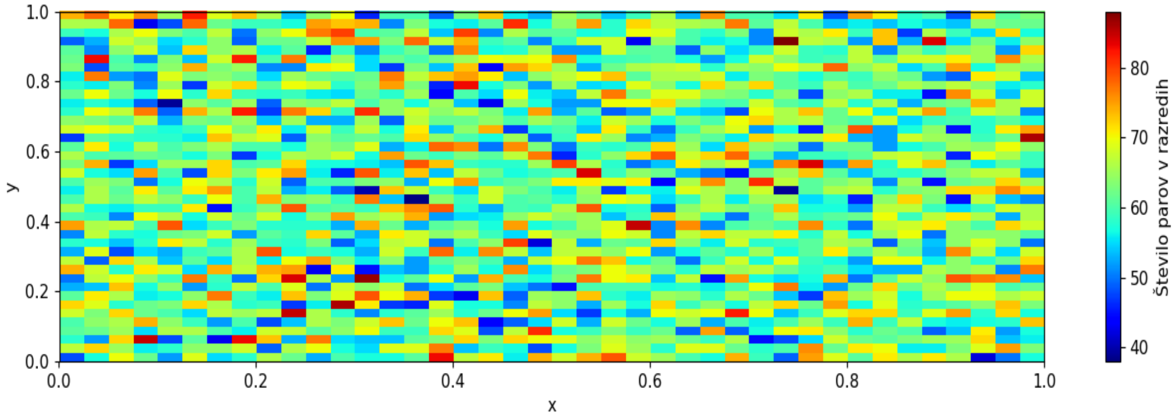
Zanima nas časovna zahtevnost generatorja Box-Muller in konvolucijskega generatorja

Slika 10: Vidimo, da je časovna zahtevnost gneratorja O(n), vendar pa je pri konvolucijskem generatorju premica približno šestkrat bolj strma.

4 Porazdelitev parov naključnih števil v prostoru

4.1 Enakomerna porazdelitev števil

Najprej vzemimo 2 para števil in se prepričajmo, da so pari števil naključno porazdeljeni po prostoru



Slika 11: Porazdelitev 1000 naključnih parov števil

Porazdelitvna funkcija je tako

$$\frac{dP}{dudv} = \frac{1}{S} = 1 \tag{5}$$

v primeru če števila enakomerno žrebamo iz kvadrata $[0,1],[0,1]$

4.2 Enakomerna porazdelitev po krogu

Sedaj si poglejmo kako bi namesto kvadrata [0,1]x[0,1] dobili enakomerno porazdeljene števile po krogu. Zapišimo verjetnost, da v krogu z radijem R izžrebamo en ločni element.

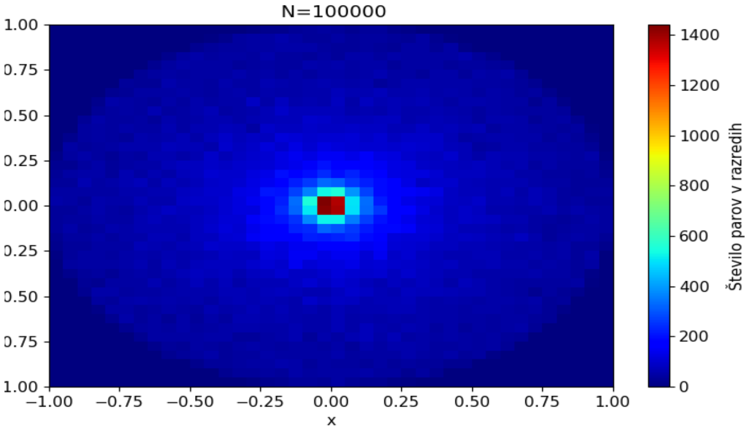
$$dP = \frac{rdrd\phi}{\pi R^2} \tag{6}$$

Ker je ta verjetnost za vse ločne elemente enaka lahko zapišemo verjetnostno porazdelitev za izžrebanje para števila na krogu na intervalu $[r, r+dr], [\phi, \phi + d\phi]$.

$$\frac{dP}{drd\phi} = \frac{r}{\pi R^2} \tag{7}$$

Vidimo, da je porazdelitev že normirana saj velja, da je $\int_0^{\inf}\int_0^{2\pi}\frac{dP}{drd\phi} = 1$. Kar opazimo je, da se verjetnost z večjimi radiji mora povečati, saj so tam ločni elementi večji.

Če naivno vzamemo dve naključni števili in zapišemo $r = u1, \phi = 2\pi u_2$. Privzeli smo, da za vse elemente dr in $d\phi$ velja enakomerna porazdelitev in zato dobimo veliko več parov števil blizu središča.



Slika 12: Porazdelitev 10000 števil po krogu, po tem ko vzamemo r = u_1 in $\phi = 2\pi u_2$, vendar pa nismo upoštevali, da ploščina ločnih elementov po krogu niso enake, kot pri enakomerni porazdelitvi temveč se večajo s radiji, zato bo tam manj števil.

Torej moramo najti pravo transformacijo iz števila $u_1, u_2 \in [0,1], [0,1]$ za kateri velja

$$\frac{dP}{dudv} = 1 \tag{8}$$

To lahko naredimo s transformacijo porazdelitev:

$$\frac{dP}{dudv} = \frac{dP}{drd\phi}|\mathbf{J}| \tag{9}$$

temu pogoju najlažje zadostimo tako, da vzamemo vsako porazdelitev posebej in zapišemo

$$\frac{dP}{dr} = \int_0^{2\pi} \frac{dP}{drd\phi} = \frac{2r}{R^2} \tag{10}$$

$$\frac{dP}{d\phi} = \int_0^{\inf} \frac{dP}{drd\phi} = \int_0^{R} \frac{dP}{drd\phi} = \frac{1}{2\pi} \tag{11}$$

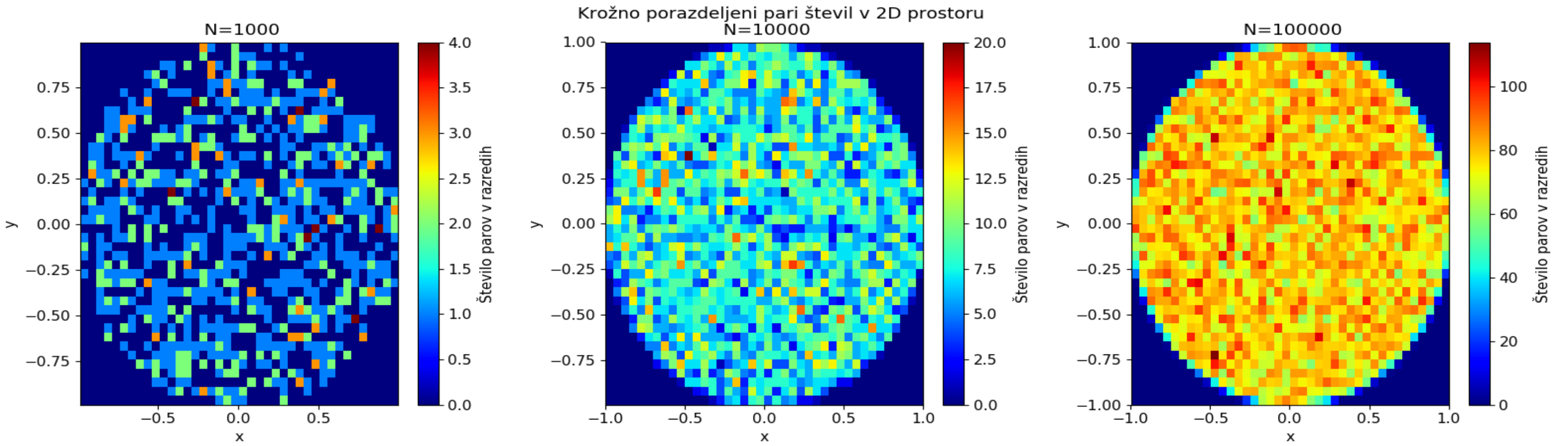
Vzemimo, da je $R = 1$, sedaj lahko pogledamo kakšen mora biti r, ϕ da bo veljalo

$$\frac{dP}{du} = \frac{dP}{dr}|\frac{dr}{du}| - > 1 = 2r|\frac{dr}{du}| - > r = \sqrt{u} \tag{12}$$

$$\frac{dP}{dv} = \frac{dP}{d\phi}|\frac{d\phi}{dv}| - > 1 = \frac{1}{2\pi}|\frac{d\phi}{du}| - > \phi = 2\phi v \tag{13}$$

Če vzamemo 2 naključni števili u_1 in u_2, lahko zapišemo r in ϕ, ki sta polarni koordinati za pare števil po krogu in sta porazdeljena enakomerno po krogu s verjetnostjo $\frac{dP}{drd\phi} = \frac{r}{\phi}$ kar je ravno enakomerna porazdelitev.

$$\begin{aligned} r &= \sqrt{u_1} \\ \phi &= 2\pi u_2 \end{aligned} \tag{14}$$



Slika 13: Enakomerna porazdelitev naključnih števil po krogu je z večanjem števil vse bolj zapolnjena. Vemo da je prava porazdelitev na disku s radijem $r = 1$ enaka $f_{R\phi}(\theta) = \frac{1}{2\pi}$

S podobno transformacijo lahko transformiramo enakomerno kotno porazdelitev, ki jo prikažemo na sferi.

4.3 Sferična enakomerna porazdelitev

Verjetnost da pri naključnem žrebu izberemo točko v krogli je enaka $dP = r^2 sin(\theta) d\phi dr d\theta = \frac{r^2 \phi dr dcos(\theta)}{4/3\pi R^3}$. Torej enakomerna porazdelitev števil po krogli z radijem $R = 1$ enaka

$$\frac{dP}{dr d\phi d\cos(\theta)} = \frac{3r^2}{4\pi} \tag{15}$$

Torej zopet vidimo, da mora biti verjetnost pri večjih radijih večja, saj so prostorski elementi večji in jih je zato manj. Prav tako pa je verjetnost po theta odvisna od $sin(theta)$, kar pomeni, so na ekvatorju prostorski elementi prav tako manjši in je zato tam verjetnost, da smo naključno izbrali tisti prostorski element večja.

Na slednjih primerih smo ugotovili, da je porazdelitev po drugih koordinatah ni konstantna tako kot v kartezičnih koordinatah, čeprav pa nam opisuje enakomerno porazdelitev elementov po krogli/krogu.

Sedaj tudi razumemo, da moramo enakomerno porazdelitev v kartezičnih koordinatah prevesti na dano porazdelitev v sferičnih koordinatah (npr. enakomerno, dipolno,...), pri tem pa moramo upoštevati Jakobijevo determinanto. Tako lahko dobimo kakšna mora biti transformacija normnalno porazdeljene spremenljivke u, da bo nova porazdelitev enaka željeni

Če nas zanima samo prostorski kot je enakomerna porazdelitev po prostorskem kotu enaka

$$\frac{dP}{d\phi d\cos(\theta)} = \frac{1}{4\pi} \tag{16}$$

zopet ločimo verjtnosti tako da pointegiramo po preostali koordinati in dobimo

$$\frac{dP}{d\phi} = \int_{-1}^{1} \frac{dP}{d\phi d\cos(\theta)} dcos(\theta) = \frac{1}{2\pi} \tag{17}$$

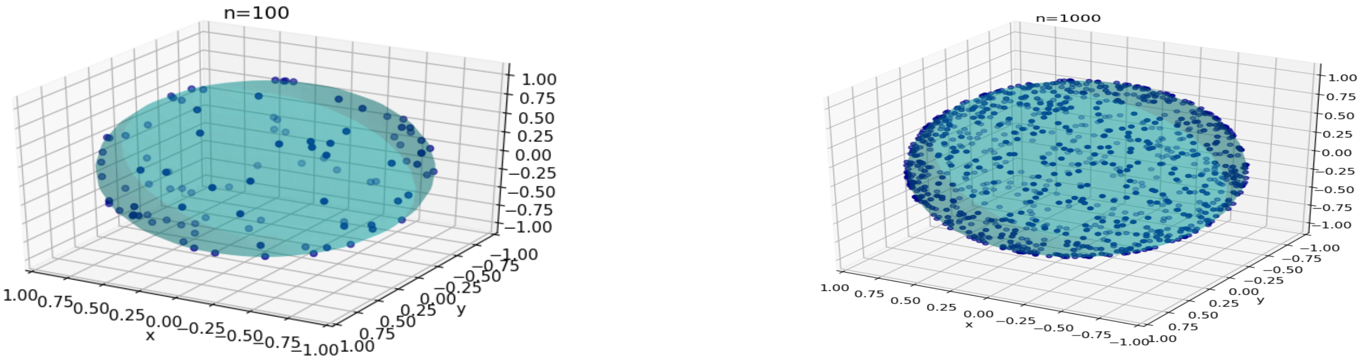
$$\frac{dP}{dcos(\theta} = \int_0^{2\pi} \frac{dP}{d\phi d\cos(\theta)} d\phi = \frac{1}{2} \tag{18}$$

Mi imamo na voljo enakomerno porazdeljeni dve števili u_1, u_2 za kateri velja $\frac{dP}{dudv} = 1$, zopet nas zanima kakšna je povezava med ϕ in u da velja za u enakomerna porazdelitev med $[0, 1]$ za ϕ pa enakomerna porazdelitev po krogu.

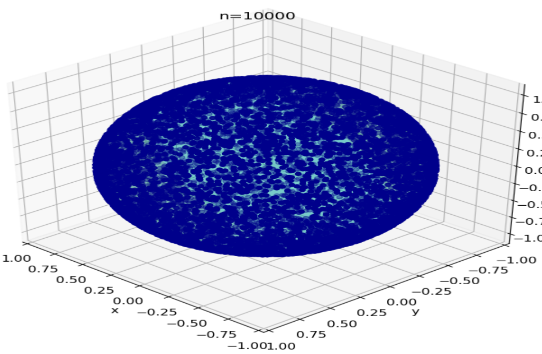
$$\mathbf{\frac{dP}{du}} = \mathbf{\frac{dP}{d\phi}} |\mathbf{\frac{d\phi}{du}}| \to 1 = \frac{1}{2\pi} |\frac{d\phi}{du}| \to \phi = 2\phi u \tag{19}$$

$$\mathbf{\frac{dP}{dv}} = \mathbf{\frac{dP}{dcos\theta}} |\mathbf{\frac{d\phi}{dv}}| \to 1 = \frac{1}{2} |\frac{dcos\theta}{dv}| \to cos\theta = 2v - 1 \tag{20}$$

konstanta $A = -1$ poskrbi za pravilmo območje $cos\theta \in [-1, 1]$, saj bi drugače prišlo območje med $[2, 0]$



Slika 14



Slika 15: vidimo da s tako transformacijo res dobimo enakomerno porazdelitev parov števil po krogli.

Poglejmo si še povprečne vrednosti različnih momentov

N	$<\theta>$	σ^2_θ	$<\phi>$	σ^2_ϕ	$\langle\cos\theta\rangle$	$\sigma^2_{\langle\cos\theta\rangle}$
100	1.5742	0.1051	3.2121	0.1215	-0.0043	0.04
1000	1.5641	0.0152	3.1321	0.0566	-0.0126	0.0043
10000	1.5712	0.0071	3.1321	0.0114	-0.0012	0.001

Vidimo, da se momenti ujemajo s teoretičnimi vrednostmi, torej povprečni $\phi = \pi$, povprečna $\theta = \frac{\pi}{2}$ in povprečni $cos\theta = 0$. Vrednosti θ se po pričakovanjih gibljejo okoli $\pi/2$, vrednosti ϕ okoli π, $\langle\cos\theta\rangle$ pa okoli ničle.

4.4 Dipolna porazdelitev naključnih števil

Sedaj pa nas namesto enakmoerne porazdelitve števil po krogli zanima kakšna druga porazdelitev, npr. porzadelitev dipolnega sevanja, ki je sorazmerna s $sin^2\theta$. Vemo, da je porazdelitev enaka

$$\frac{dP}{d\Omega} = Asin^2(\theta) = \frac{3}{8\pi}sin^2(\theta) \tag{21}$$

Izverednotimo sedaj porazdelitev po vsakem kotu posebej.

$$\begin{aligned} \frac{dP}{dcos\theta} &= \frac{3}{4}sin^2(\theta) \\ \frac{dP}{d\phi} &= \frac{1}{2\pi} \end{aligned} \tag{22}$$

Zanima nas zopet kako zagotoviti, da enakomerna porazdelitev med $[0, 1]$ po u pri transformaciji iz $u \to \theta$ pridelala dipolno porazdelitev , tako da za porazdelitveno funkcijo, da velja $\frac{dP}{du} = \frac{dP}{dcos\theta}|\frac{d\cos\theta}{du}|$.

$$1 = \frac{3}{4}sin^2(\theta)|\frac{d\cos\theta}{du}| \tag{23}$$

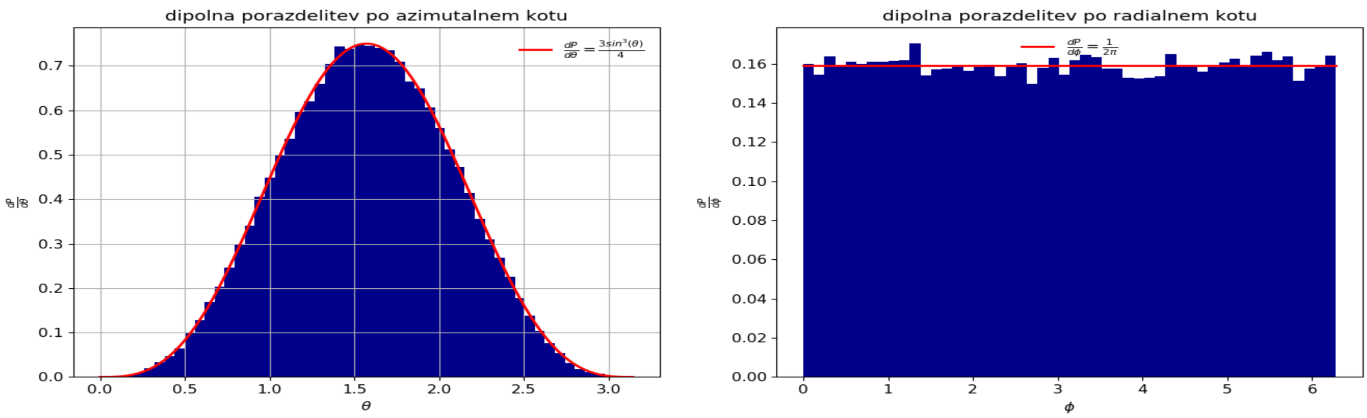
$$du = \frac{3}{4}sin^2(\theta)dcos\theta \tag{24}$$

$$u = \frac{1}{4}(3cos\theta - cos^3\theta) + \frac{1}{2} \tag{25}$$

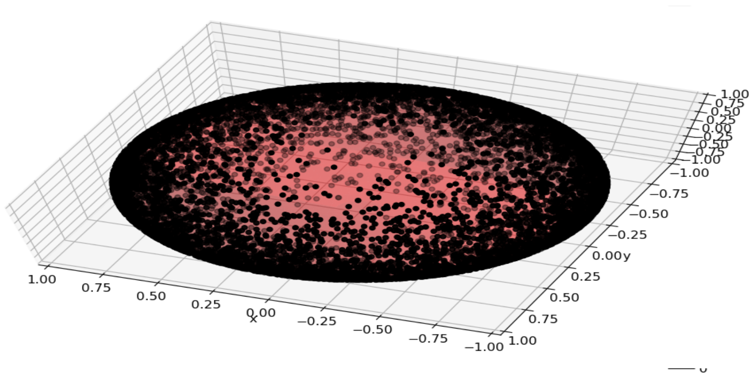
$A = \frac{1}{2}$ da velja $u \in [0, 1]$. Te enačbe ne moremo obrniti zato moramo problem rešiti numerično. Za vsak u moramo najti ustrezen $\cos\theta$. Če bomo sedaj tako naključno izžrebali števila in izračunali vsakič $\cos\theta$, ki zadošča enačbi (25) bi morala naša porazdelitev ustrezati dipolnem sevanju. Po ϕ pa je transformacija enaka kot pri enakomerni porazdelitvi pri krogli, torej $\phi = 2\pi v$.

4.5 Numerično reševanje transformacijske enačbe

Enačbo (25) rešujmo numerično s iteracijskimi metodami (newtnova, sekantna), bisekcijo, ali pa uporabimo $python fsolve$. Sam sem uporabil slednjo in za $N = 10000$ naključnih števil med $[0, 1]$ izračunal ustrezne θ.



Slika 16: Porazdelitev dipolnega sevanja po kotu ϕ in θ



Slika 17: Porazdelitev na krogli vidimo, da dipol seva le v eni smeri. V našem primeru torej sumimo, da je dipol usmerjen v smeri z, saj je sevanje pravokotno iz njega najmočnejše.

Poglejmo si še momente in jih preverimo s teoretičnimi

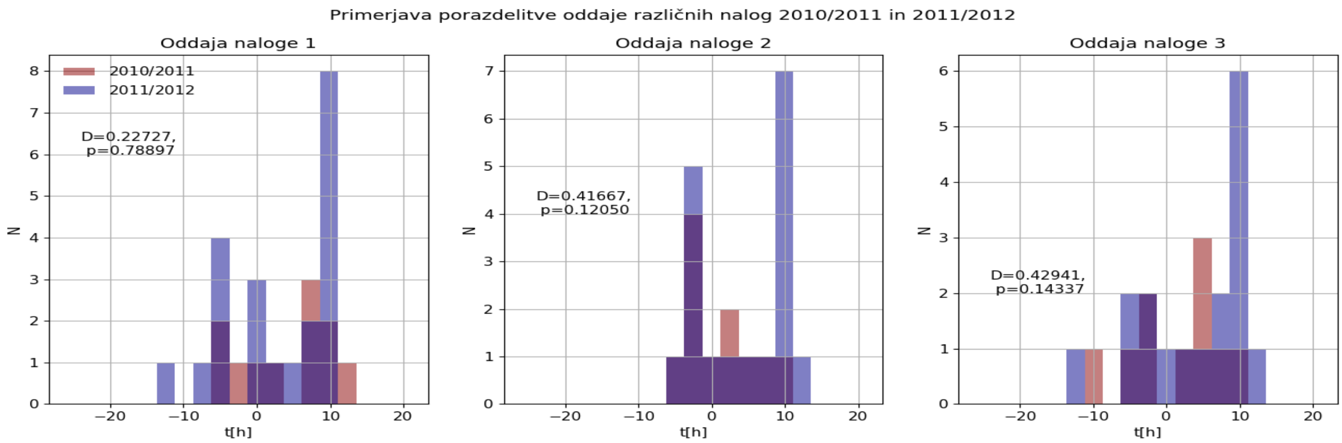
Tabela 1: Vrednosti povprečnega θ, ϕ, $\cos\theta$, $\cos^2\theta$ in njihovih varianc za različne N

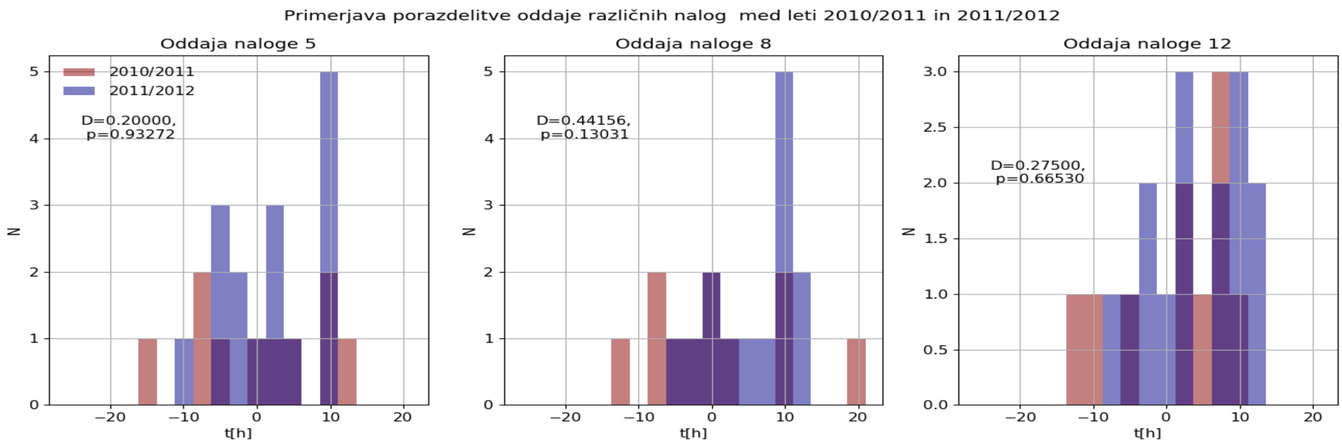
N	$<\theta>$	σ^2_θ	$<\phi>$	σ^2_ϕ	$\langle\cos\theta\rangle$	$\sigma^2_{\langle\cos\theta\rangle}$	$\langle\cos^2\theta\rangle$	$\sigma^2_{\langle\cos^2\theta\rangle}$
1000	1.5821	0.011	3.15345	0.0744	-0.0007	0.0014	0.1769	0.0033
10000	1.5677	0.0023	3.1413	0.0112	0.0011	0.0004	0.2041	0.0012

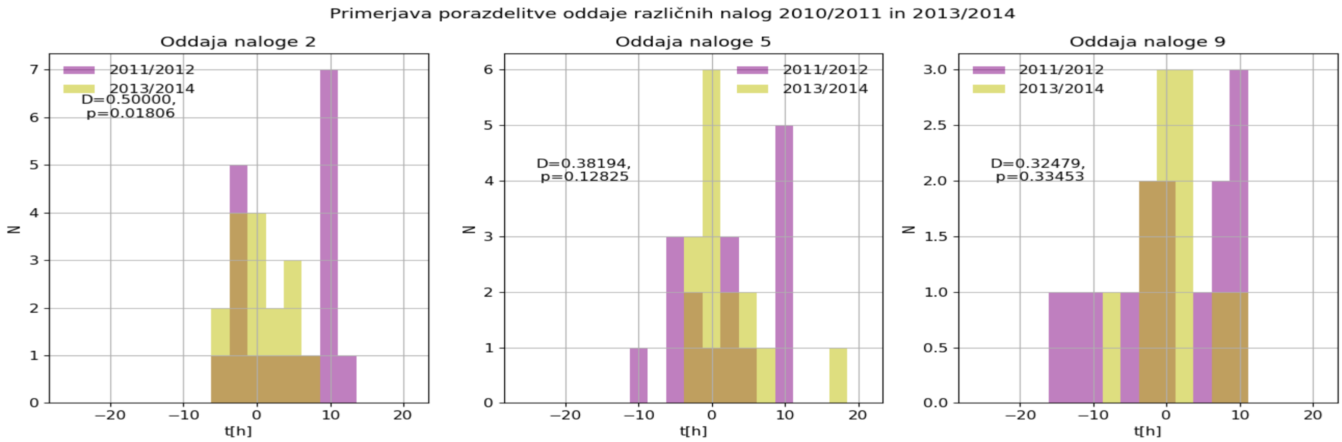
Tudi tu momenti ustrezajo teoretičnim.

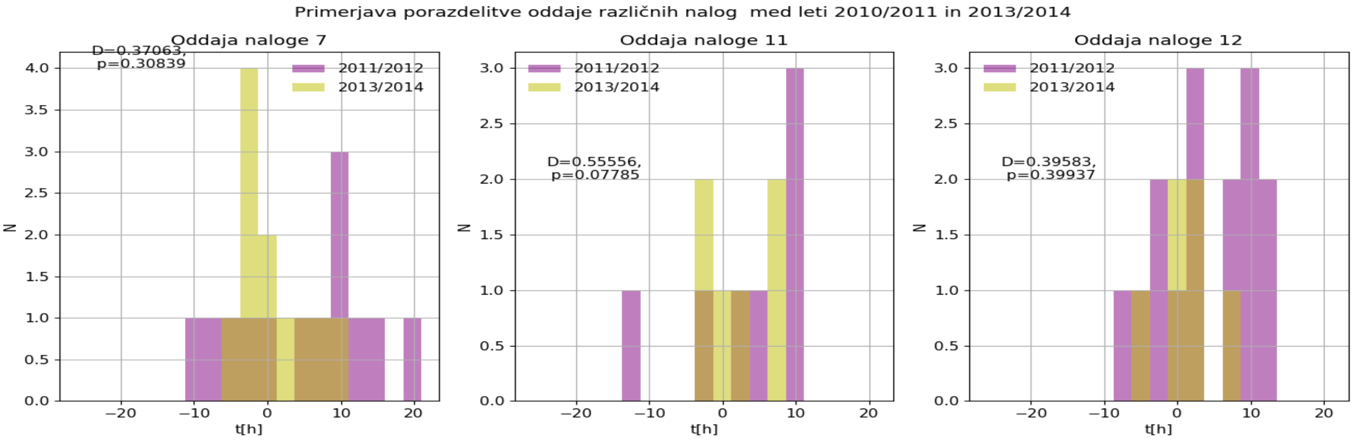
5 Statistika oddaje domačih nalog iz Modelske Analize I

Na voljo imamo podatke o oddaji modelskih nalog za leta 2010/11,2011/12,2012/13,2014/15. Iz nje lahko naredimo statistično obdelavo in najdemo razlike med letniki in nalogami. Poglejmo si najprej razlike v oddaji med generacijami.


Primerjava porazdelitve oddaje različnih nalog 2010/2011 in 2011/2012
Oddaja naloge 1
2010/2011
2011/2012
D=0.22727,
p=0.78897
Oddaja naloge 2
D=0.41667,
p=0.12050
Oddaja naloge 3
D=0.42941,
p=0.14337
N
t[h]

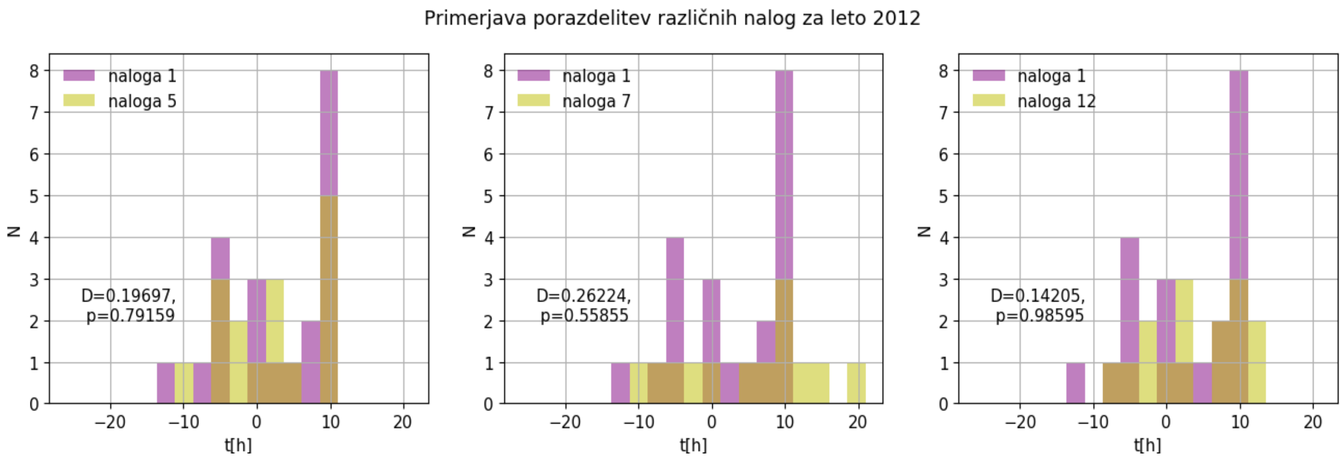

Primerjava porazdelitve oddaje različnih nalog med leti 2010/2011 in 2011/2012
Oddaja naloge 5
2010/2011
2011/2012
D=0.20000,
p=0.93272
Oddaja naloge 8
D=0.44156,
p=0.13031
Oddaja naloge 12
D=0.27500,
p=0.66530
N
t[h]

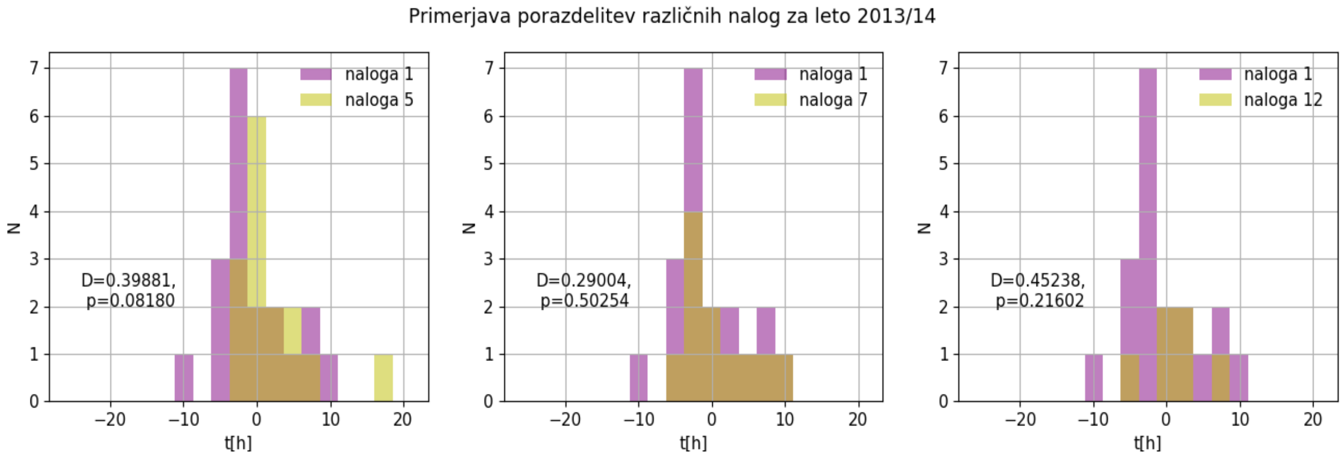

Primerjava porazdelitve oddaje različnih nalog 2010/2011 in 2013/2014
Oddaja naloge 2
2011/2012
2013/2014
D=0.50000,
p=0.01806
Oddaja naloge 5
2011/2012
2013/2014
D=0.38194,
p=0.12825
Oddaja naloge 9
2011/2012
2013/2014
D=0.32479,
p=0.33453
N
t[h]



Enakost porazdelitev testiramo s testom Kolmogorov - Smirnov, in ugotovimo da razlike v oddaji nalog so (npr 11 naloga med letoma 2010/11 in 2013/14 ter 2 naloga med letom 2010/2013), kar lahko tudi vidimo, saj je bila generacija 2013/14 veliko bolj pridna, kot generacija 2010/12.

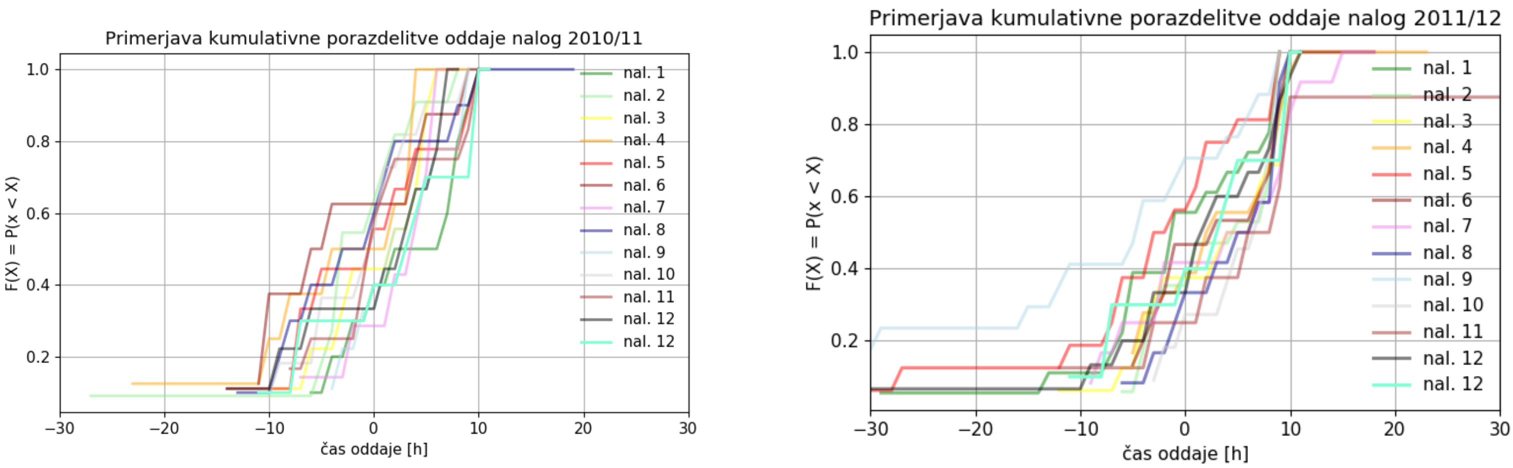
5.1 Primerjava oddaj različnih nalog znotraj iste generacije





Slika 18: vidimo, da se znotraj generacije porazdelitve nalog razlikujejo, prav tako se močno razlikuje število, vendar pa v danih primerih ne moremo govoriti o statistično pomembni razliki med porazdlitvami naloge 1,5 ; 1,7 in 1,12. Verjetno je med kakim drugim datummom sigurno statistična pomembna razlika

Zanimivi so grafi kumulativne porazdelitve oddaje nalog



Slika 19: Vidimo, je ob času četrtek 00:00 oddana le polovica vseh nalog. Iz slednjega grafa se da razbrati tudi "outlinerje", kot npr. zadnja naloga 12. pri letu 2010/11 in npr. naloga 7, katero je veliko ljudi ponoči nekaj ur prepozno in npr. naloga 11 generacije 2011/12.

6 Zaključek

Pogledali smo si kako delujejo generatorji naključnih števil in kako iz enakomerne porazdelitve dobiti katero koli porazdelitev. Statistične porazdelitve smo testitrali s testi.