



*Univerza v Ljubljani*  
*Fakulteta za matematiko in fiziko*

*Oddelek za fiziko*

---

## 8. naloga: - Generatorji naključnih števil

---

POROČILO PRI PREDMETU MODELSKA ANALIZA 1

2015/2016

*Avtor:*

Klemen RAHNE  
28152028

20. november 2016

# 1 Teorija-uvod

V tokratni nalogi bomo preverjali generatorje naključnih števil. Generatorje bomo preverili s testom Kolmogorov-Smirnova ter testom  $\chi^2$ .

## 1.1 Test $\chi^2$

Pri testiranju porazdelitve števil v testu  $\chi^2$ , najprej razpon števk razdelimo v  $m$  predalov. V  $i$ -tem predalčku je torej  $N_i$  števk, oz. verjetnost da posamezna številka pade v  $i$ -ti predalček je  $\rho' = \frac{N_i}{n}$ , kjer je  $n$  število vseh števk. Pri vsakem testu moramo tudi predpostaviti analitično porazdelitev. Analitična verjetnost, da pade številka v  $i$ -ti predalček je:

$$\rho_i = \int_{x_i}^{x_{i+1}} \frac{dP}{dx} dx \quad (1.1)$$

kjer je  $\frac{dP}{dx}$  analitična porazdelitev,  $x_i, x_{i+1}$ , spodnja in zgornja meja predala. Test  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^m \frac{(N_i - \rho_i n)^2}{n \rho_i} \quad (1.2)$$

Ker imamo veliko število  $n$ , je v limiti  $\chi^2$  porazdeljen neodvisno od porazdelitve  $\frac{dP}{dx}$ :

$$\frac{dP}{d\chi^2} = \frac{1}{2^{\frac{m-1}{2}} \Gamma(m - \frac{1}{2})} (\chi^2)^{\frac{m-3}{2}} e^{-\frac{\chi^2}{2}} \quad (1.3)$$

Lahko določimo tudi zgornjo mejo  $\chi^2_{>}$ , ki na stopnjo tveganja  $\alpha$  dovoljuje testni  $\chi^2$ :

$$\int_{\chi^2_{>}}^{\infty} \frac{dP}{d\chi^2} d\chi^2 = \alpha \quad (1.4)$$

Vrednosti  $\chi^2_{>}$  za posamezno stopnjo tveganja  $\alpha$  so že tabelirane.

## 1.2 Test Kolmogorov-Smirnov

Prav tako test Kolmogorov-Smirnov testira izmerjeno porazdelitev z analitično. Tokrat je potrebno imeti obe porazdelitve v kumulativni obliki:

$$F(x) = \int_{-\infty}^x \frac{dP}{dx} dx \quad (1.5)$$
$$F_i = \frac{N_i}{N}$$

kjer je  $N_i$  število vseh števk v porazdelitvi manjših od  $x_i$ , katera je zgornja meja  $i$ -tega predalčka.  $N$  je število vseh števk v porazdelitvi. Pri testu Kolmogorov-Smirnov testiramo vrednost:

$$D = \sup |F_i - F(x_i)| \quad (1.6)$$

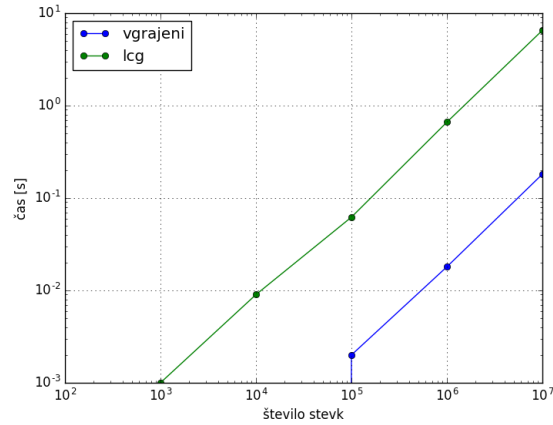
ki predstavlja največje odstopanje testne kumulativne porazdelitve od analitične porazdelitve. Ponovno je  $D$  porazdeljen ter lahko ponovno na stopnji tveganja  $\alpha$  določimo zgornjo vrednost  $D_{>}$ :

$$\lim_{n \rightarrow \infty} P(D\sqrt{N} < D_{>}) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 D_{>}^2} = 1 - \alpha \quad (1.7)$$

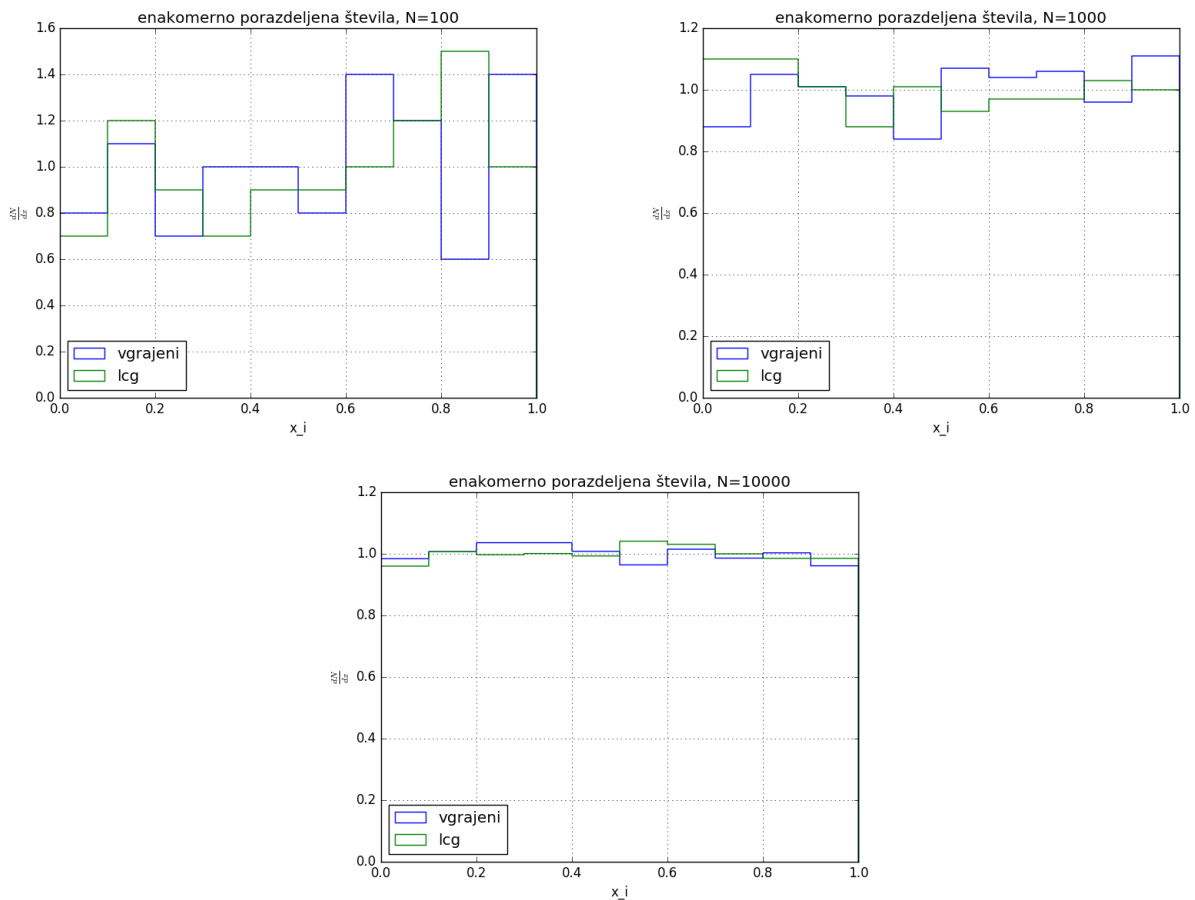
Tudi vrednosti  $D_{>}$  so tabelirane.

# 2 Generatorji enakomerno porazdeljenih števil

Oglejmo si generator naključnih števil. Primerjali bomo dva generatorja. Prvi bo vgrajeni generator naključnih števil v programskem jeziku *Python*, ki je osnovan na algoritmu Mersenne Twister. Drugi generator je linearni kongruenčni generator, z  $a = 1103515245$ ,  $b = 123454$  in  $m = 2^{32}$ .



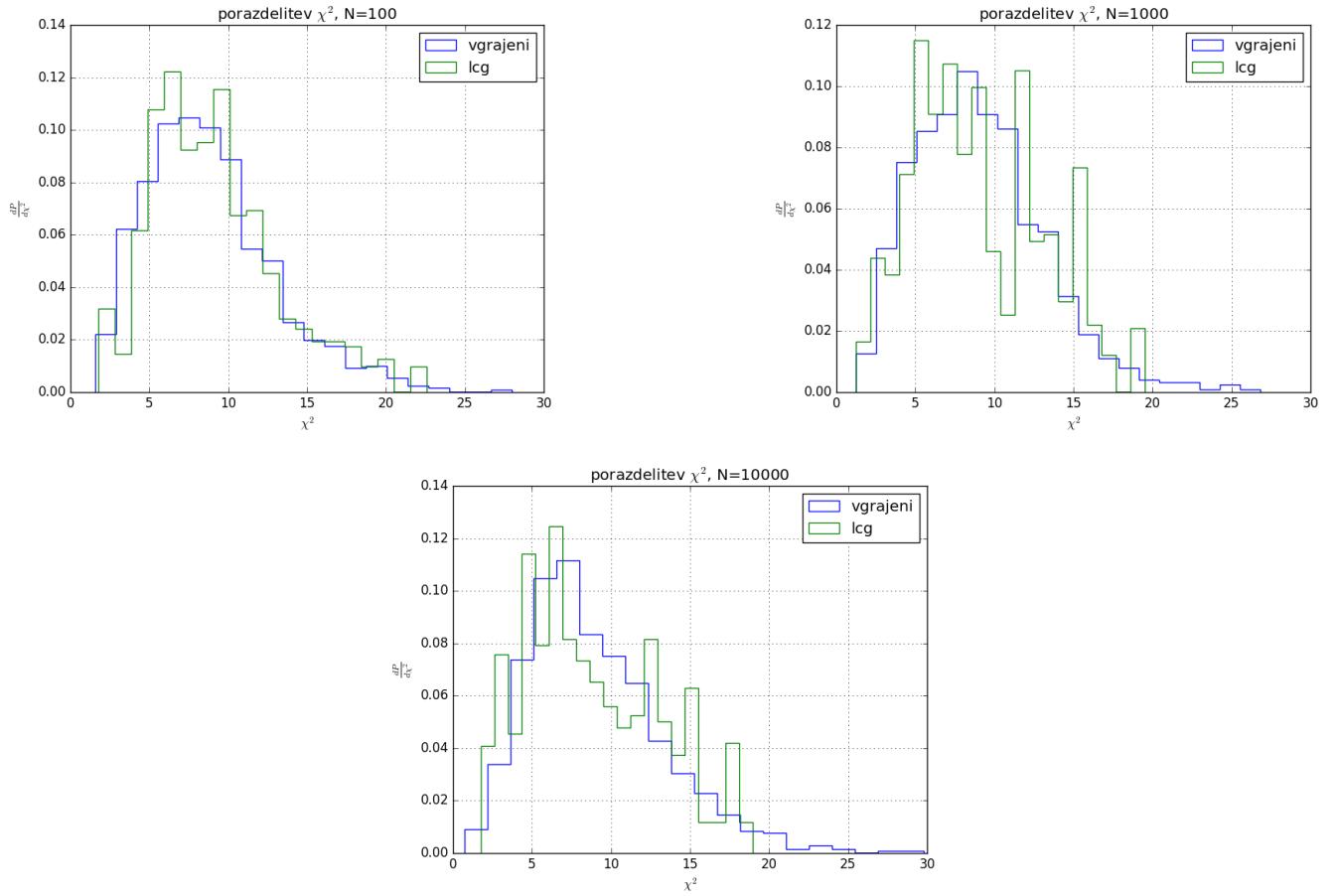
Slika 2.1: Primerjava hitrosti testiranih generatorjev. Opazimo, da z naraščanjem potrebnih naključnih števil, skupni čas narašča linearno oz. za eno naključno številko potrebujemo enako časa, neodvisno od števila potrebnih števil. Opazimo, da je vgrajeni za več kot en velikostni razred hitrejši od linearnega. To je verjetno posledica tega, ker smo vgrajeni generator klicali iz knjižnice *Numpy*, katera je specializirana za matematične operacije. Funkcijo za linearni kongruenčni generator sem napisal sam.



Slika 2.2: Primerjava dveh generatorjev naključnih števil. Na prvi pogled ne moremo oceniti, njune pravilnosti/napačnosti delovanja. Opazimo, da z naraščanjem števila  $N$ , se porazdelitev bliža analitični vrednosti, ki je 1.

## 2.1 Test $\chi^2$

Testirajmo sedaj zgornja generatorja s testom  $\chi^2$  in pogledjmo njune porazdelitve vrednosti  $\chi^2$ .



Slika 2.3: Porazdelitev  $\chi^2$  obeh generatorjev v odvisnosti od števila generiranih števil. Oblike sta podobni, le da *lcg* generator ima nekoliko večjo variacijo, od idealne zvezne porazdelitve.

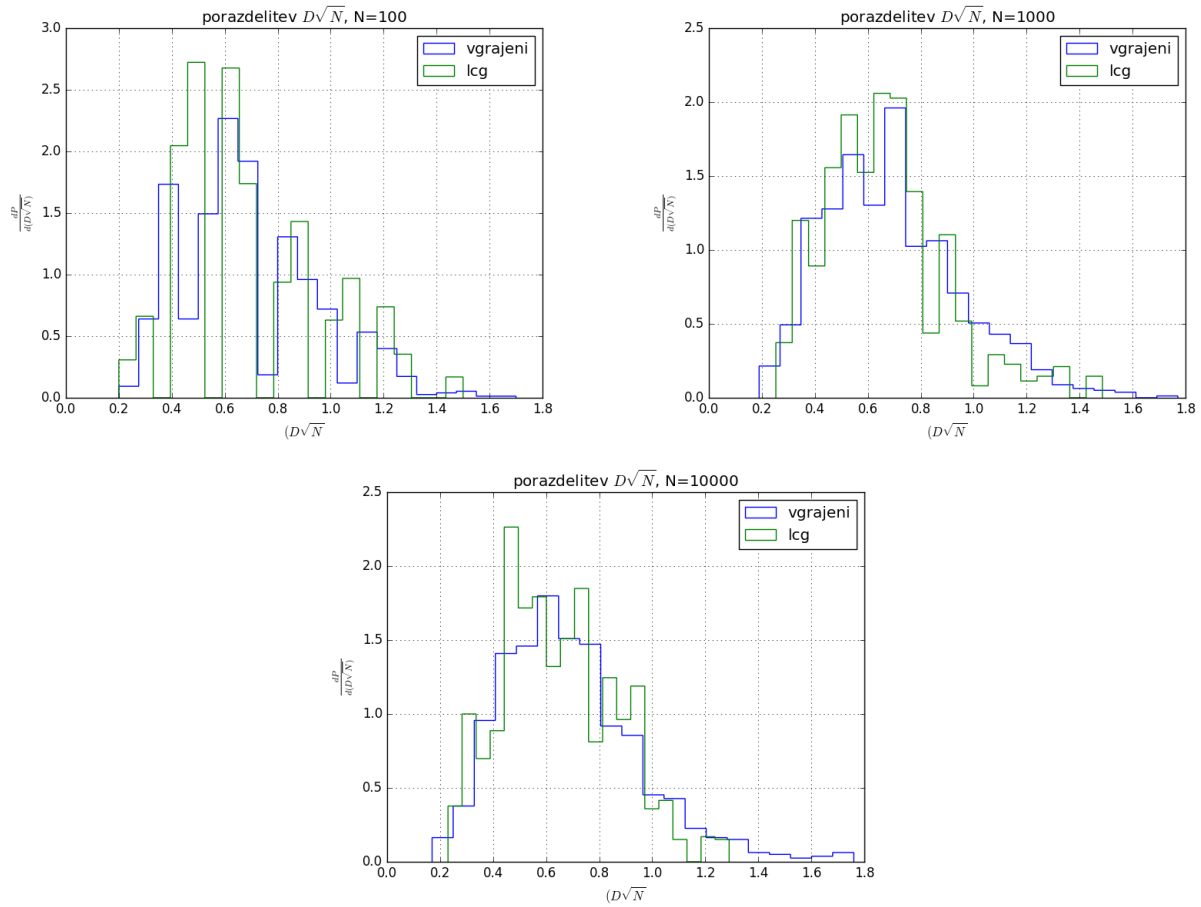
Primerjajmo to porazdelitev še številčno. Iz enačbe 1.4 in njenih vrednosti lahko primerjamo naša generatorja naključnih števil. Za enakomerno porazdeljena števila so pričakovane vrednosti  $\chi^2$ :

stopnja tveganja $\alpha$	0.01	0.5	0.99
$m = 9, \chi^2 :$	21.67	8.34	2.09

generator	<i>lcg</i>		vgrajeni	
$N$	povprečni $\chi^2$	variacija $\chi^2$	povprečni $\chi^2$	variacija $\chi^2$
100	9.05	4.35	9.34	4.23
1000	9.18	4.23	9.04	4.09
10000	9.17	4.41	9.05	4.27

Tabela 2.1: Izmerjene vrednosti  $\chi^2$ . Med generatorjema ni opaznih razlik v tem testu. Oba generatorja imata povprečno vrednost, ki je v okolici mejne vrednosti za  $\alpha = 50\%$ . Potrebno pa je tudi opozoriti na zelo veliko varianco vrednosti  $\chi^2$ .

## 2.2 Test Kolmogorov-Smirnov



Slika 2.4: Porazdelitev  $\chi^2$  obeh generatorjev v odvisnosti od števila generiranih števil. Oblike sta podobni, le da ima *lcg* generator nekaj izrazitih vrhov.

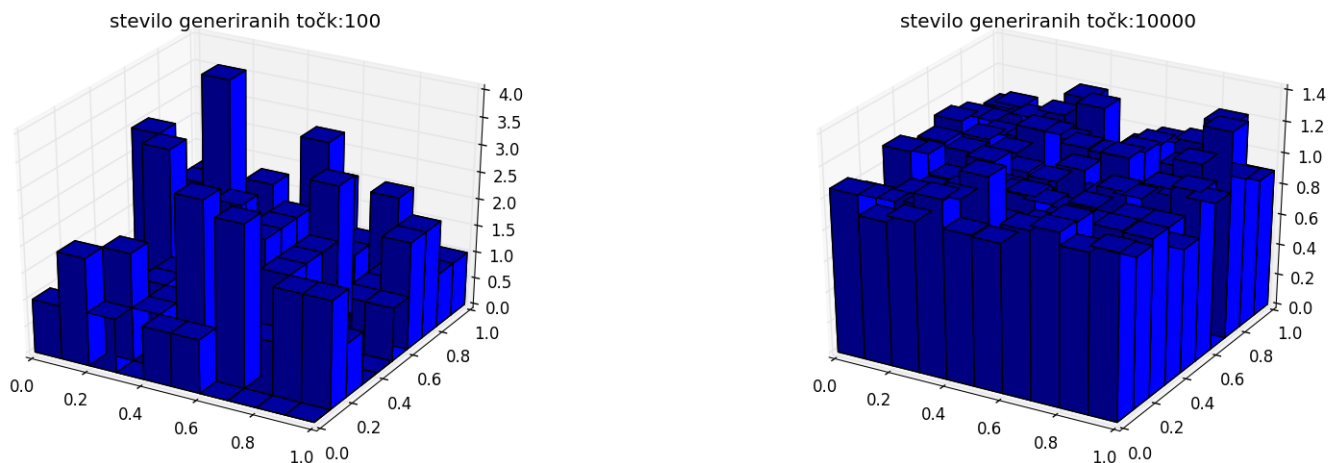
Iz enačbe 1.7 dobimo za naš primer naslednje mejne vrednosti  $D\sqrt{N}$ :

stopnja tveganja $\alpha$	0.01	0.5	0.99
$D\sqrt{N}$	0.44	0.83	1.63

generator	<i>lcg</i>		vgrajeni	
$N$	povprečni $D\sqrt{N}$	variacija $D\sqrt{N}$	povprečni $D\sqrt{N}$	variacija $D\sqrt{N}$
100	0.68	0.26	0.70	0.28
1000	0.70	0.26	0.66	0.24
10000	0.69	0.27	0.66	0.23

Tabela 2.2: Tako, kot pri testu  $\chi^2$ , imamo tokrat vrednosti  $D\sqrt{N}$ , porazdeljene okoli mejne vrednosti za stopnjo tveganja  $\alpha = 50\%$ . Tudi tokrat je variacija pričakovane vrednosti velika. Vzrok velike variacije (ist lahko sklepamo za veliko variacijo pri testu  $\chi^2$ ) in velike 50% stopnje tveganja nam, pove, da sta generatorja sprejemljiva in dejansko generirata naključne številke.

## 2.3 Dvodimenzionalna porazdelitev



Slika 2.5: Porazdelitev naključnih števil v dveh dimenzijah. Ponovno se lepo vidi, da z večanjem generiranih števil, se porazdelitev približuje konstantni vrednosti.

Kaj pa nam pove test  $\chi^2$ . Sedaj imamo  $10 * 10$  točk, ki jih primerjamo, oz  $100 - 1$  prostostnih stopenj. Iz tabel sledijo mejne vrednosti  $\chi^2$ :

stopnja tveganja $\alpha$	0.01	0.5	0.99
$\chi^2$	69.23	98.33	134.64

$N$	povprečni $\chi^2$	variacija $\chi^2$
1000	99.01	14.06
10000	99.05	14.11
100000	98.29	14.54

Tabela 2.3: Ponovno rezultati testa  $\chi^2$  niso presenetljivi. Tudi tokrat povprečna vrednost  $\chi^2$  zavzema mejno vrednost pri pričakovani vrednosti 50%. Tokrat so variacije približno štirikrat manjše.

## 3 Sevanje v prostoru

### 3.1 Naključna smer v prostoru

Smer v prostoru najpreprosteje opišemo s sfernima kotoma  $\phi$  in  $\theta$ . Za prehod iz kartezičnih v sferične koordinate moramo upoštevati še "Jacobija":

$$\frac{dP}{dx dy dz} = \frac{dP}{dr d\theta d\phi} r^2 \sin \theta = \text{const.} \quad (3.1)$$

Ker nas zanima samo smer izsevanih fotonov, lahko radialno komponento zanemarimo. Z upoštevanjem tega, ter normalizacijo verjetnostne funkcije po prostoru dobimo:

$$\frac{d^2 P}{d(\cos \theta) d\phi} = \frac{1}{4\pi} \quad (3.2)$$

Kot  $\phi$  zavzema vrednosti med 0 in  $2\pi$ , ter zaradi preproste zveze v zgornji enačbi vidimo:

$$\frac{dP}{d\phi} = \frac{1}{2\pi} \quad (3.3)$$

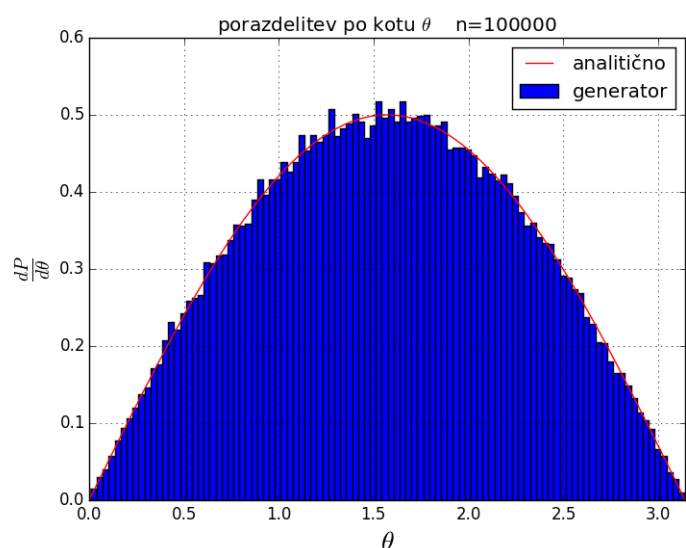
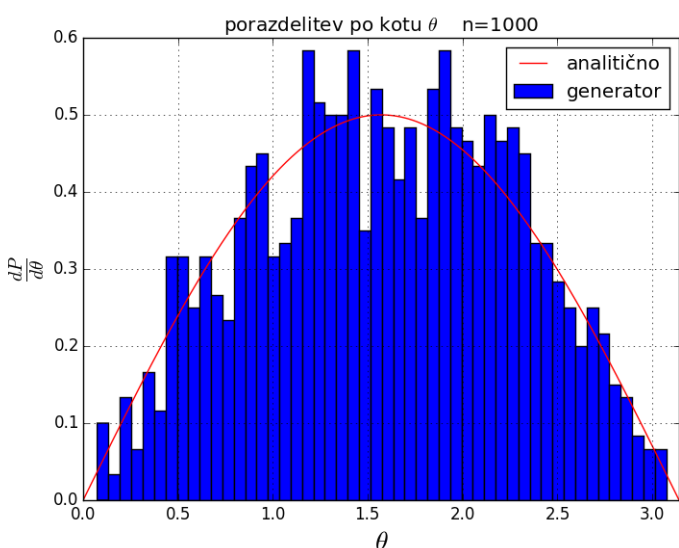
Za kot  $\theta$  potem velja:

$$\frac{dP}{d(\cos \theta)} = \frac{1}{2} \quad (3.4)$$

Iz zgornjih enačb imamo naslednji porazdelitvi kotov:

$$\begin{aligned} \phi &= 2\pi u \\ \theta &= \arccos(2v - 1) \end{aligned} \quad (3.5)$$

kjer sta  $u, v$  porazdeljena enakomerno med 0 in 1. Porazdelitev po kotu  $\phi$  je trivialna (enakomerna med 0 in  $2\pi$ ), zato si oglejmo porazdelitev po kotu  $\theta$ .



Slika 3.1: Porazdelitev kota  $\theta$ , pri naključni smeri v prostori. Na levem grafu imamo 1000 naključnih smeri in 50 razdelčnov, na desni pa  $10^5$  naključnih smeri ter 100 razdelčkov. Rdeča krivulja, je analitična rešitev. Z dovolj veliko številom naključnih smeri se izbrane smeri približujejo analitični rešitvi.

Primerjajmo kvaliteto našega generatorja naključnih smeri številčno. Za primerjavo vzemimo računanje nekaj osnovnih momentov. Moment se matematično izračuna:

$$\langle M \rangle = A \int_0^\pi d\theta \int_0^{2\pi} d\phi M(\theta, \phi) \sin \theta d\theta d\phi \quad (3.6)$$

kjer je  $M$  funkcija našega momenta ter  $A$  normalizacijska konstanta. V našem primeru je normalizacijska konstanta enaka  $\frac{1}{4\pi}$ . Za numerični izračun momenta funkcije, se zgornja enačba preoblikuje v:

$$\langle M \rangle = \frac{1}{N} \sum_{i=1}^N M(x_i) \rho(x_i) \quad (3.7)$$

kjer je  $M$  funkcija momenta, ter  $\rho$  normirana porazdelitvena funkcija.

moment	analitična vrednost	generator
$\langle \theta \rangle$	$\pi/2$	1.7530
$\langle \cos \theta \rangle$	0	-0.0017
$\langle \phi \rangle$	$\pi = 3.1417$	3.1561
$\langle \cos \phi \rangle$	0	-0.0016
$\langle \sin \theta \rangle$	$\pi/4 = 0.7854$	0.7848
$\langle Y_2^0 \rangle$	0	0.0013
$\langle \cos^2 \theta \rangle$	1/3	0.3338
$\langle \cos^2 \phi \rangle$	1/5	0.4992

### 3.2 Sevanje dipola

Fotoni so pri dipolnem sevanju prostorsko porazdeljeni:

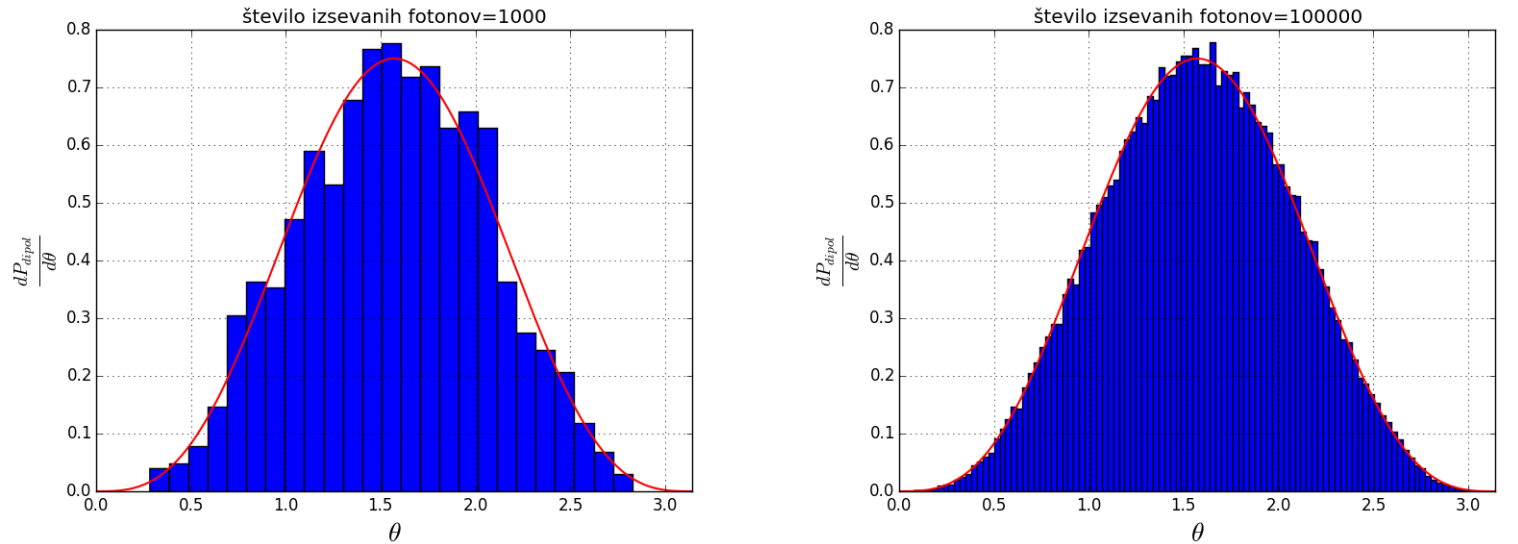
$$\frac{dP}{d\Omega} = A \sin^2 \theta \quad (3.8)$$

Iz normalizacije sledi vrednost normalizacijske konstante  $A = \frac{3}{8\pi}$ . Trivialno se vidi, da je kot  $\phi$  enakomerno porazdeljen med 0 in  $2\pi$ . Za določitev porazdelitve po kotu  $\theta$ , najprej zgornjo porazdelitev transformiramo v kumulativno porazdelitev:

$$F(\theta) = u = \frac{3}{8\pi} \int_0^{2\pi} \int_0^\theta \sin^2 \theta' \sin \theta' d\theta' d\phi' \quad (3.9)$$

$$u = \frac{3}{8\pi} \left[ \frac{\cos^3 \theta}{3} - \cos \theta + \frac{2}{3} \right]$$

Spremenljivka  $u$  je sedaj enakomerno porazdeljena med 0 in 1, ter sedaj poiščemo inverz zgornje funkcije oz. izpostavimo  $\theta$ . Ker tega ne moremo rešiti analitično, bomo inverz poiskali s pomočjo bisekcije.



Slika 3.2: Porazdelitev smeri izsevanih fotonov po kotu  $\theta$  (za kot  $\phi$  je porazdelitev enakomerna med 0 in  $2\pi$ ). Na levem grafu imamo 1000 fotonov in 25 razdelčnov, na desni pa  $10^5$  fotonov ter 100 razdelčkov. Rdeča krivulja, je analitična rešitev:  $\frac{dP_{dipol}}{d\theta} = \frac{3}{4} \sin^3 \theta$ .

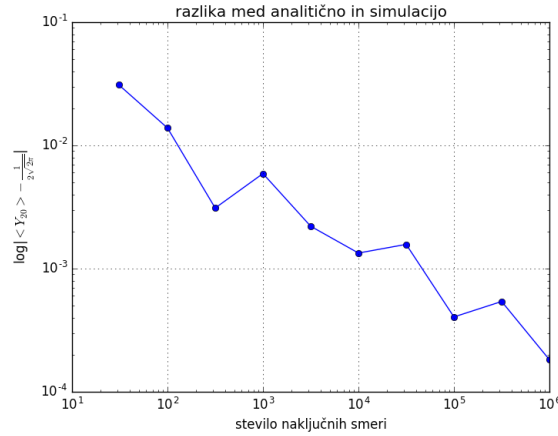
Tako kot v prejšnjem poglavju si oglejmo nekaj osnovnih momentov:

Oglejmo si kako hitro konvergirajo izračunane vrednosti k analitični vrednosti.



moment	analitična vrednost	generator
$\langle \theta \rangle$	$\pi/2 = 1.5708$	1.5729
$\langle \cos \theta \rangle$	0	-0.0020
$\langle \phi \rangle$	$\pi = 3.1417$	3.1561
$\langle \cos \phi \rangle$	0	-0.0016
$\langle \sin \theta \rangle$	$\frac{9}{32}\pi = 0.8836$	0.8840
$\langle Y_2^0 \rangle$	$\frac{1}{2\sqrt{2\pi}} = 0.1262$	-0.1267
$\langle \cos^2 \theta \rangle$	1/5	0.1995
$\langle \cos^2 \phi \rangle$	1/2	0.4992

Tabela 3.1: Za računanje momentov smo uporabili  $10^5$  fotonov in 100 predalčkov pri risanju histograma.



Slika 3.3: Hitrost približevanja analitični vrednosti. Na obeh skalah imamo logaritemsko skalo in lahko sklepamo, da se s povečevanjem števila števok oz. fotonov linearno povečujemo veljavna mesta izračunani vrednosti.

## 4 Box-Mullerjev generator in konvolucijski generator

V prvi nalogi smo testirali enakomerno porazdeljene naključne številke. Sedaj bomo testirali dva generatorja naključnih števil, ki so porazdeljene po Gaussu. Prvi generator je *Box\_mullerjev* generator, ki generira števke  $x$  z uporabo enakomerno porazdeljene naključne spremenljivke  $x_1$  in  $x_2$  ter enačbe:

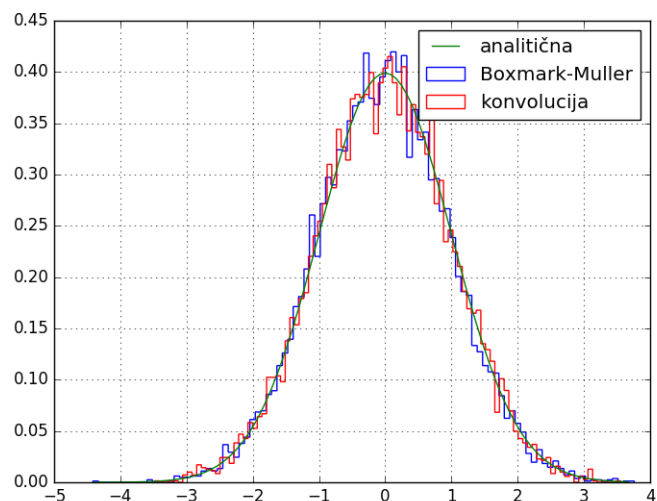
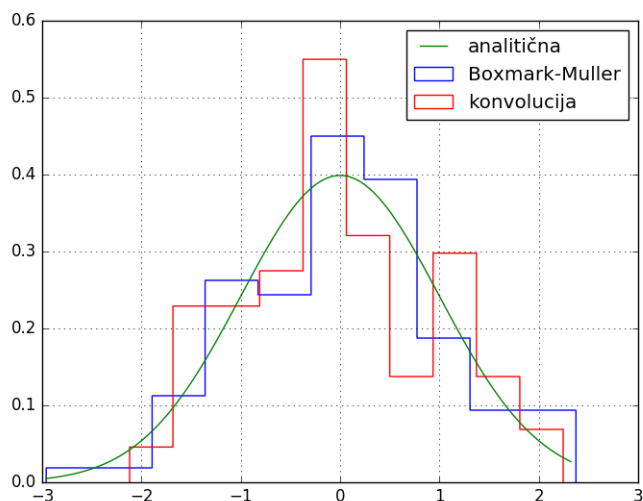
$$\frac{dP}{dx} = \sqrt{-2 \log x_1} \cos(2\pi x_2) \quad (4.1)$$

Drugi generator se naslanja na centralni limitni izrek, ki pravi, da je vsota  $N$  enakomerno porazdeljenih naključnih spremenljivk enakih normalni porazdelitvi. V našem generatorju bo  $N = 12$ , ker pa se vrh premakne za  $\frac{N}{2}$ , ga je potrebno še prej prestaviti v izhodišče. Širino predpostavimo, da je neodvisna od in je enaka 1. za večje  $N$ -je je širina odvisna od  $N$ .

Še mejne vrednosti za testa  $\chi^2$  in test Kolmogorov-Smirnov:

$\alpha$	0.01	0.5	0.99
$m = 49 \chi^2$	28.9	48.3	74.9
Kolmogorov-Smirnov	0.44	0.83	1.63

Iz testiranja opazimo, da je Box-Mullerjev test boljši od konvolucijskega. Lahko da k temu rezultatu vpliva dejstvo, da pri konvoluciji imamo območje števok med 0 in 12, medtem ko ima prava porazdelitev vrednosti tudi izven tega intervala.



Slika 4.1: Primerjava porazdelitev različnih generatorjev normalno porazdeljenih števil. Na desni imamo 10000 generiranih števil in 100 predalčkov, medtem, ko je na levi strani 100 števk in 10 predalčkov. Inženirsko lahko rečemo, da generatorja generirata dobro porazdelitev števil.

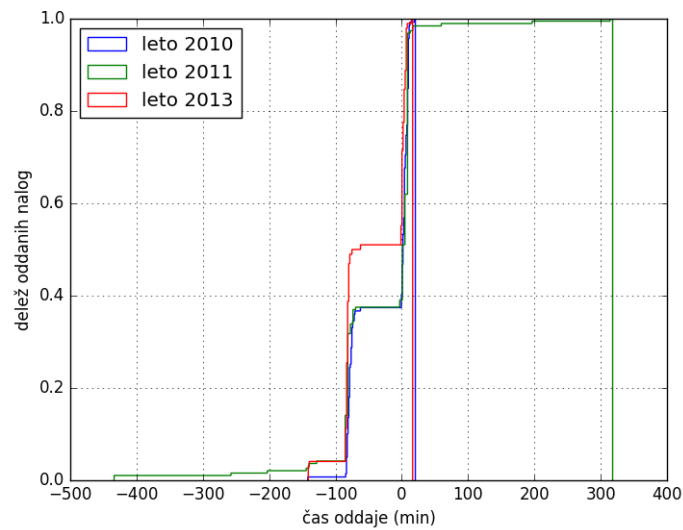
generator	Konvolucija		Box-Muller	
$N$	povprečni $\chi^2$	variacija $\chi^2$	povprečni $\chi^2$	variacija $\chi^2$
1000	46.88	12.26	59.96	16.68
10000	49.56	10.62	62.67	51.61
100000	83.37	13.45	62.23	32.45

generator	konvolucija		Box-Muller	
$N$	povprečni $D\sqrt{N}$	variacija $D\sqrt{N}$	povprečni $D\sqrt{N}$	variacija $D\sqrt{N}$
1000	0.77	0.27	0.76	0.27
10000	0.81	0.24	0.71	0.20
100000	1.23	0.23	0.69	0.25

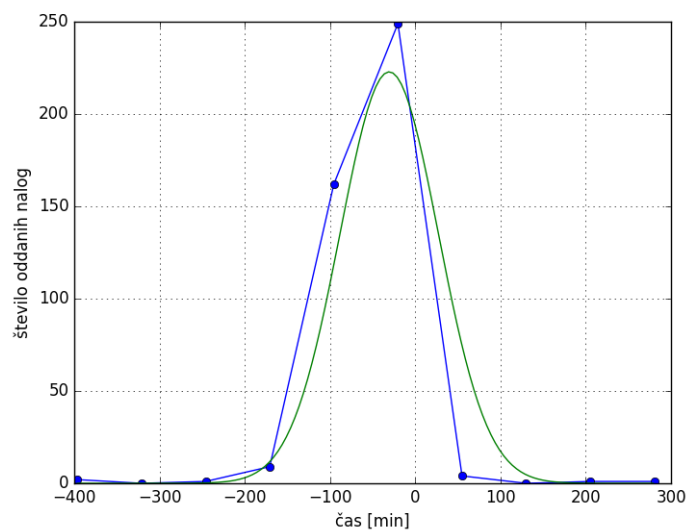
## 5 Časi oddaj nalog

Poglejmo si še kako so v preteklih letih oddajali domače naloge pri predmetu modelska analiza. Iz kumulativne porazdelitve oddaj nalog opazimo, da je večina nalog že oddanih pred rokom. Opazimo tudi dve izrazite stopnici pri vseh letih, ena pri okoli  $-100min$ , druga pa pri okoli  $0min$ .

Ker je funkcija  $erf(x)$  v grobem približku podobna stopnici, predpostavimo, da je čas oddaje porazdeljen Gaussovsko. Testirali bomo vsa leta skupaj, predpostavimo, da so vsa leta oddajali pod enakimi pogoji. Najprej iz dane porazdelitve prilagajamo normalno funkcijo, da dobimo vrh in širino časa oddajanja. Dobimo: čas  $-30.5min$  in širino  $57.7min$ .



Slika 5.1: Kumulativna porazdelitev oddanih nalog po posameznih letih.



Slika 5.2: POrazdelitev oddanih nalog in njihova Gaussova prilagoditev. Kljub temu, da je prilagajanje izgleda dobro, testa  $\chi^2$  in Kolmogorov ne prestane, saj so vrednosti občutno prevelike:  $\chi^2 = 9217901$  in  $D\sqrt{N} = 3.54$ .