



Modelska Analiza 1

8. naloga - Generatorji slučajnih števil

Avtor: Matic Lubej
Asistent: dr. Simon Čopar
Predavatelj: prof. dr. Alojz Kodre

Ljubljana, 4.12.2013

Naloga:

Pri tej nalogi smo se spoznali z različnimi generatorji naključnih števil. V prvem delu smo generatorje testirali s statističnimi testi kot sta test χ^2 in test Kolmogorov-Smirnova in tako preverili njihovo naključnost. V drugem delu naloge smo generirali naključne kote in preverili, če se skladajo s enakomerno porazdeljenim prostorskim kotom in v primeru dipolnega sevanja z ustrezno porazdelitvijo za dipol. Tretji del naloge je bil precej podoben prvemu, le da smo namesto enakomerno porazdeljenih števil testirali kakovost generatorjev normalno porazdeljenih števil. V zadnjem delu naloge smo kot zanimivos pogledali kumulativne porazdelitve oddajnih časov nalog pri predmetu Modelska analiza I in jih med seboj statistično primerjali. Nalogo sem reševal z uporabo matematičnega orodja Mathematica.

Del I

Teorija

V tem delu bom najprej predstavil teorijo, ki je bila uporabljena v kasnejših delih naloge. S teorijo je predvsem mišljen test χ^2 in pa test Kolmogorov-Smirnova.

1 Test χ^2

V primeru naključnih števil, si izberemo interval od 0 do 1. Ta interval razdelimo na S razredov, v katerih ležijo slučajna števila z_1, z_2, \dots, z_N . Verjetnost p_k , da število leži v k -tem razredu, ki ga določata meji razreda x_{k-1} in x_k , je enaka:

$$p_k = \int_{x_{k-1}}^{x_k} \frac{dP}{dx} dx. \quad (1)$$

Če imamo uniformno porazdelitev, velja:

$$\frac{dP}{dx} = \text{konst.} \quad (2)$$

Ker imamo N slučajnih števil, pričakujemo, da bo v k -ti razred padlo Np_k števil, kjer bo v našem primeru veljalo $p_k = 1/S$, razlike $N_k - N/S$ pa bi morale biti statistično majhne, kar lahko preverimo s testom χ^2 :

$$\chi^2 = \sum_{k=1}^S \frac{(N_k - N/S)^2}{N/S}. \quad (3)$$

V limiti, ko imamo opravka z velikim številom N , je porazdelitev vrednosti χ^2 neodvisna od porazdelitvene funkcije števil in je enaka:

$$\frac{dP}{d\chi^2} = \frac{1}{2^{(n-1)/2} \Gamma(n-1/2)} (\chi^2)^{(n-3)/2} e^{-\chi^2/2}; \quad \chi^2 \geq 0. \quad (4)$$

Določimo lahko zgornjo mejo, tako da velja:

$$\int_{\chi^2_0}^{\infty} \frac{dP}{d\chi^2} d\chi^2 = \alpha, \quad (5)$$

kjer je α stopnja tveganja, na podlagi katere lahko potrdimo ali zavržemo ujemanje vzorca s predpisano porazdelitvijo. Omenimo lahko še to, da ima porazdelitev tudi analitično izračunljivo povprečno vrednost in pa varianco, ki sta enaki:

$$\langle \chi^2 \rangle = \nu - 1, \quad (6)$$

$$\langle \chi^2 \rangle^2 = 2\nu - 2, \quad (7)$$

kjer je ν število prostostnih stopenj porazdelitve. To nam daje še dodaten test, s katerim lahko preverimo pravilnost porazdelitve.

2 Test Kolmogorov-Smirnova

Test K-S uporabimo takrat, ko imamo na voljo veliko število točk in želimo preveriti, če podpirajo privzeto zvezno porazdelitveno funkcijo. Formuliran je za zvezno kumulativno porazdelitveno funkcijo $F(z)$, definirano kot:

$$F(z) = \int_{-\infty}^z \frac{dP}{dx} dx. \quad (8)$$

Sestavimo lahko še eksperimentalno porazdelitveno funkcijo $f(z)$ kot:

$$f(z) = \frac{k(z)}{N}, \quad (9)$$

kjer je $k(z)$ število slučajnih števil, ki so manjša od z . Funkcija $F(z)$ je naraščajoča in gladka, $f(z)$ pa naraščajoča in stopničasta. Obe funkciji sta normirani, kar pomeni, da velja $F(\infty) = f(\infty) = 1$. Pri testiranju ujemanja je odločujoča absolutna vrednost razlike med obema funkcijama:

$$D = \sup_{-\infty < z < \infty} |f(z) - F(z)|. \quad (10)$$

Če so slučajna števila v skladu s porazdelitvenim zakonom, je ta razlika le izjemoma zelo velika. Kolmogorov je pokazal, da velja:

$$\lim_{n \rightarrow \infty} P(D\sqrt{N} < d) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 d^2}. \quad (11)$$

Pri testiranju si izberemo mejni $d_>$, tako da je verjetnost, da bo pri danem vzorcu $D\sqrt{N}$ manjši od $d_>$ enaka $1 - \alpha$. Če je pri kakem vzorcu $D\sqrt{N}$ večji od dovoljene meje, potem lahko s tveganjem α zavrnemo hipotezo.

Del II

Enakomerna porazdelitev slučajnih števil

1 Naloga

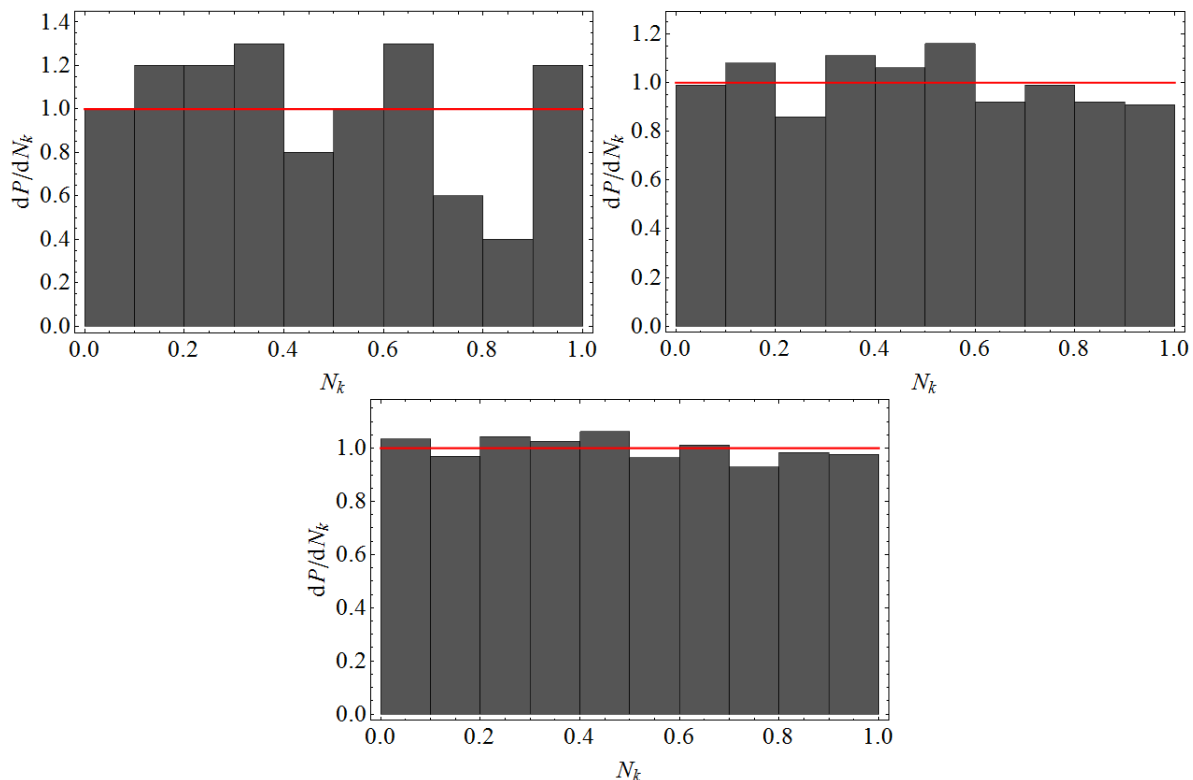
Preveriti smo morali pravilnost enakomernih porazdelitev različnih generatorjev z različnimi testi za različno dimensionalne porazdelitve. V naslednjih razdelkih so predstavljeni testi χ^2 in K-S testi za različne generatorje naključnih števil. Na koncu smo preverili še časovno učinkovitost generatorjev. Uporabil sem generatorje, ki mi jih je nudilo matematično orodje Mathematica, ki so:

- Mathematica privzeti generator
- Kongruenčni generator
- Intelov MKL generator
- Generator Mersenne Twister

2 1D porazdelitev

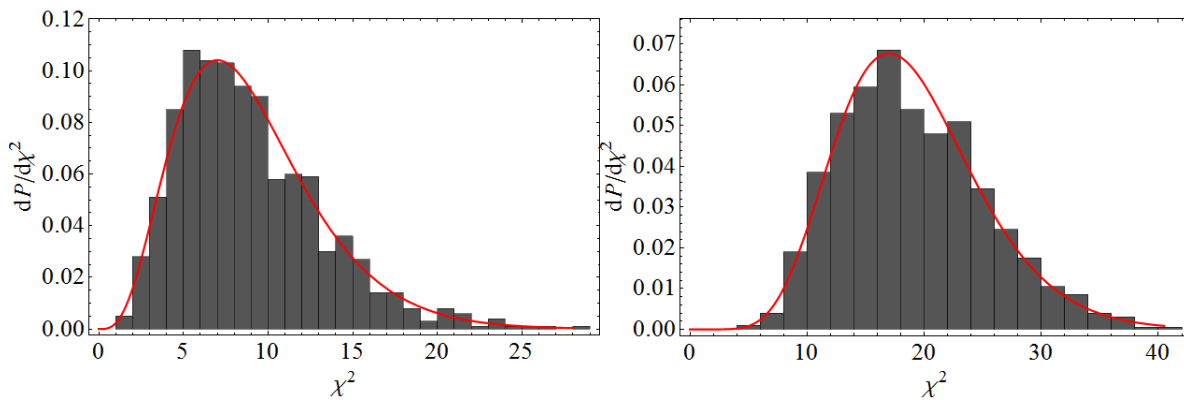
2.1 1D Test χ^2

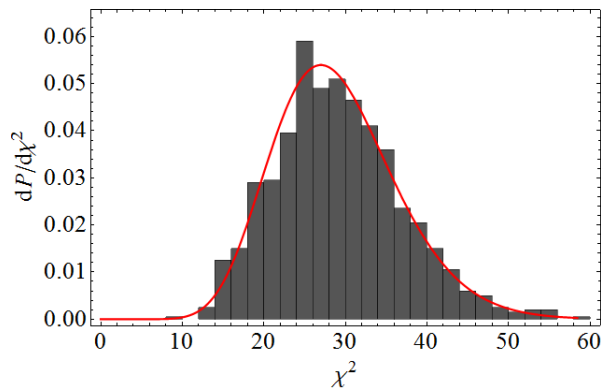
V glavnem se vsi ti generatorji množično uporabljajo po svetu, zato na prvi pogled nobeden ni nič boljši od drugega. Oglejmo si histogram porazdelitev za 100, 1000 in 10 000 naključnih števil:



Edina razlika med temi porazdelitvami je natančnost, koliko se ujema s teoretično, kar pa je odvisna od števila števil. Več kot imamo naključnih števil, bolj je porazdelitev podobna teoretični.

Za takšno porazdelitev lahko izvednotimo χ^2 , izračun pa nato ponovimo M -krat. Za $M = 10\,000$, sem izračunal porazdelitev χ^2 in ga predstavil v obliki histograma v verjetnostni obliki ter ga primerjal s teorijsko krivuljo. Spodaj so prikazane porazdelitve za različne vrednosti N , spet samo za eno metodo, ker so si med seboj porazdelitve precej podobne:

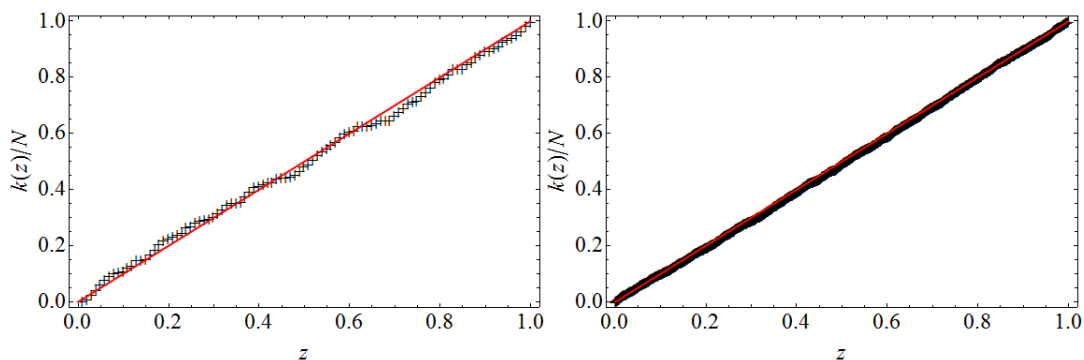




Zgornje porazdelitve so bile izračunane za $N = 100$ in $S = 10$, $N = 1000$ in $S = 20$ ter za $N = 10\,000$ in $S = 30$. Vidimo, da se histogrami lepo prilegajo porazdelitvam χ^2 , hkrati pa opazimo, da se za veliko število razredov S porazdelitev približuje normalni. Rezultati testov so prikazani na koncu razdelka.

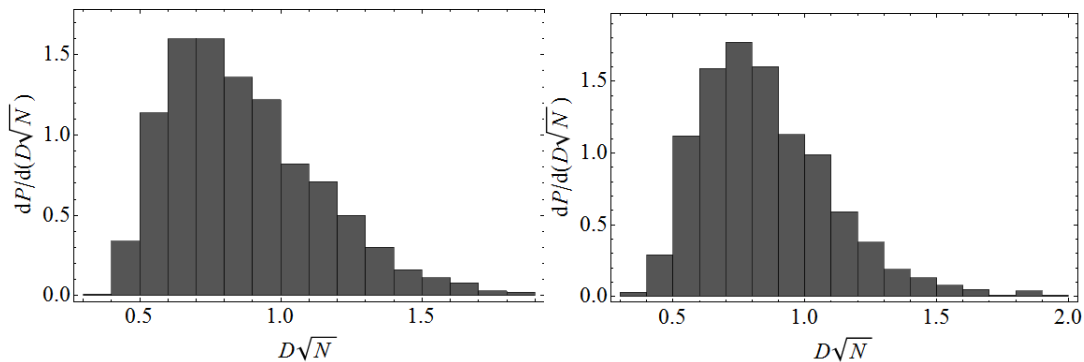
2.2 Test K-S

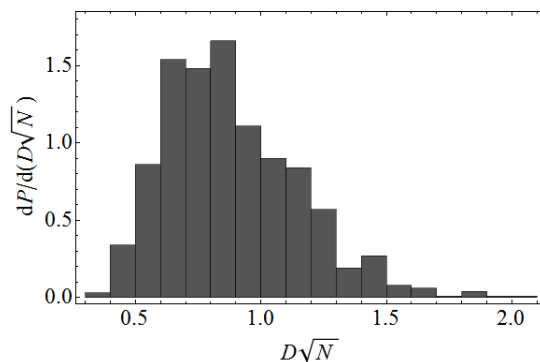
Podobno kot v prejšnjem podrazdelku, lahko sedaj napravimo še test Kolmogorov-Smirnova, pri katerem pa potrebujemo kumulativno porazdelitev naključnih števil, kot je opisano na začetku poročila. Tudi v tem primeru sem naredil test za kumulativno porazdelitev za 100, 1000 in 10 000 naključnih števil:



Zgoraj sta za boljše razumevanje prikazana le primera za $N = 100$ in 1000. Vidimo, da se kumulativna porazdelitev lepo prilega premici, torej pričakujemo dobre rezultate.

Za takšno porazdelitev lahko iz vrednotimo $D\sqrt{N}$, ki je omenjen na začetku pri teoriji, izračun pa nato ponovimo M -krat. Za $M = 10000$, sem izračunal porazdelitev $D\sqrt{N}$ in ga predstavil v obliki histograma v verjetnostni obliki. Spodaj so prikazane porazdelitve za različne vrednosti N , tudi v tem primeru samo za eno metodo, ker so si med seboj porazdelitve precej podobne:





Vidimo, da so si sicer precej podobne tudi med seboj, saj nimamo takšnih odstopanj kot pri razredih pri testu χ^2 . Vidimo, da porazdelitev v splošnem leži med 0.4 in 1.8, kaj pa to pomeni za sprejem oziroma zavrnitev hipoteze, si pa oglejmo v naslednjem podrazdelku.

2.3 Rezultati testov

Da bomo lahko med seboj primerjali rezultate različnih testov za porazdelitve iz različnih metod, si najprej oglejmo tabelo relevantnih vrednosti testov χ^2 in K-S za različne stopnje tveganja α :

| | | | | | | | |
|---------------|------|----------------|----------------|----------------|-------------|------|----------|
| | P | $\nu = 10 - 1$ | $\nu = 20 - 1$ | $\nu = 30 - 1$ | | P | |
| $\chi^2_{d>}$ | 0.01 | 21.67 | 36.19 | 50.89 | K-S $_{d>}$ | 0.01 | 0.441028 |
| | 0.5 | 8.34 | 18.34 | 29.34 | | 0.5 | 0.827574 |
| | 0.99 | 2.09 | 7.63 | 14.95 | | 0.99 | 1.62762 |

Sedaj, ko imamo napisano tabelo za primerjalne vrednosti, si oglejmo izračunane vrednosti, ki nam jih da posamezen generator:

| | | | | | |
|-----------|------------------|-----------------------|------------------|---------------------|------------|
| $N = 100$ | Metoda | Test: 9 ± 4.24264 | Metoda | Test: 19 ± 6.16 | $N = 1000$ |
| $S = 10$ | Default | 9.00 ± 4.19 | Default | 18.75 ± 6.13 | $S = 20$ |
| | Kongruenčni | 8.97 ± 4.38 | Kongruenčni | 18.83 ± 6.07 | |
| | Intel MKL | 8.82 ± 4.11 | Intel MKL | 18.75 ± 6.29 | |
| | Mersenne Twister | 8.91 ± 4.23 | Mersenne Twister | 19.09 ± 6.17 | |
| | Metoda | Test: 29 ± 7.62 | | | |
| | Default | 29.02 ± 7.76 | | | |
| | Kongruenčni | 29.14 ± 7.81 | | | |
| | Intel MKL | 29.08 ± 7.66 | | | |
| | Mersenne Twister | 29.61 ± 7.83 | | | |
| | $N = 10000$ | | $S = 30$ | | |

Pri testih χ^2 smo predstavili 4 različne metode pri treh različnih množicah naključnih števil z različnimi števili razredov. Vsaka tabela prikazuje rezultate metod za različno velike množice naključnih števil, kjer je v prvi vrstici tudi navedena tudi teoretična pričakovana vrednost in standardni odklon. Vidimo, da praktično ni razlike med metodami, niti ne med različnimi množicami naključnih števil. Vse metode opravijo test, saj imajo porazdelitve χ^2 prav takšne vrednosti, kot jih napoveduje teorija. Opazimo, da χ^2 skače po kar velikem intervalu, s čimer pokrije celotno verjetnostno območje s povprečjem $\alpha \approx 50\%$. To, da vrednost stopnje tveganja α skače, ne pomeni, da je naša metoda slaba, ampak ravno to potrjuje naključnost podatkov.

Oglejmo si še rezultate testa K-S:

| $N = 100$ | Metoda | Test K-S | Metoda | Test K-S | $N = 1000$ |
|------------------|-----------------|------------------|-----------------|-----------------|------------|
| | Default | 0.84 ± 0.25 | Default | 0.86 ± 0.26 | |
| Kongruenčni | 0.86 ± 0.27 | Kongruenčni | 0.86 ± 0.26 | | |
| Intel MKL | 0.85 ± 0.25 | Intel MKL | 0.87 ± 0.26 | | |
| Mersenne Twister | 0.86 ± 0.26 | Mersenne Twister | 0.85 ± 0.26 | | |

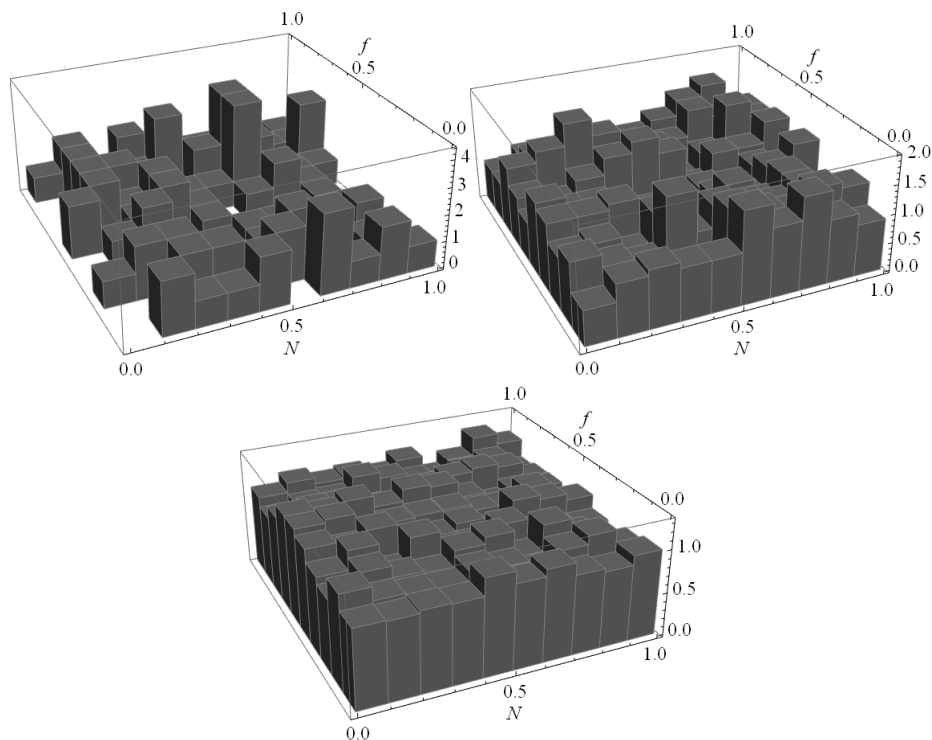
| Metoda | Test K-S |
|------------------|-----------------|
| Default | 0.86 ± 0.25 |
| Kongruenčni | 0.86 ± 0.26 |
| Intel MKL | 0.87 ± 0.26 |
| Mersenne Twister | 0.86 ± 0.25 |
| $N = 10000$ | |

V tem primeru vidimo, da so rezultati testa v vseh primerih enaki. Spet ni praktične razlike med metodami, rezultati K-S testa pa vodijo k istemu zaključku kot pri testu χ^2 , saj vrednost testa skače po kar velikem intervalu in tako zavzame celotno verjetnostno območje, povprečje pa ima pri $\alpha \approx 50\%$. Tudi pri tem testu ta podatek potrjuje naključnost dogodkov, rezultate pa zaključimo s sklepom, da so vse metode kvalitetne.

3 2D porazdelitev

3.1 Test χ^2

Dober generator naključnih števil ni samo ta, ki naključno generira posamezna števila, ampak tudi zaporedne pare števil, trojice in tako dalje do n -te dimenzije. V tem razdelku bomo preverili naključnost 2D porazdelitve parov naključnih števil, testirali pa ga bomo na podoben način kot prej, in sicer s testom χ^2 . Za začetek si oglejmo histograme porazdelitev 2D naključnih števil za $N = 100, 1000$ in 10000 :

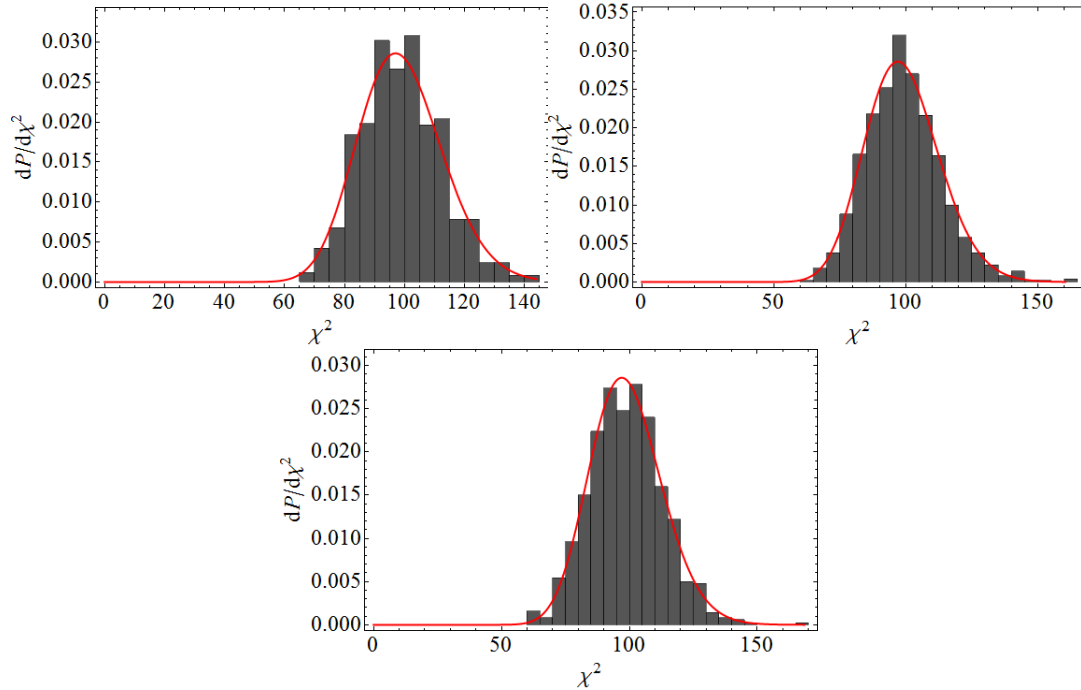


Vidimo, da porazdelitve za manjše vrednosti N niso preveč enakomerne, saj je količina števil pre-majhna glede na število razredov. Porazdelitev se izboljšuje, ko gremo proti večjim vrednostim N . Tudi za takšno porazdelitev lahko izračunamo χ^2 , kjer izračun ponovimo $\times 10\,000$. Čeprav je porazdelitev dvodimenzionalna, lahko vseeno χ^2 porazdelitev predstavimo z navadnim histogramom, če eksperimentalne izmerke primerjamo z vrednostjo N/s^2 in nato seštejemo po vseh razredih. Teoretična pričakovana vrednost takšne porazdelitve je tedaj:

$$\langle \chi^2 \rangle = \nu^2 - 1, \quad (12)$$

$$\langle \chi^2 \rangle^2 = 2\nu^2 - 2. \quad (13)$$

Spodaj so prikazane porazdelitve χ^2 za zgornje tri primere:



Porazdelitve so si v tem primeru precej podobne, zato sklepamo, da so vse metode sprejemljive tudi v primeru 2D porazdelitev. Zaradi večjega števila razredov je sedaj porazdelitev tudi bolj podobna normalni porazdelitvi.

3.2 Rezultati testa

Še za primer 2D porazdelitev lahko sedaj primerjamo kritične vrednosti χ^2 z eksperimentalnimi. Spodnja tabela prikazuje tabelo relevantnih vrednosti za različne stopnje tveganja α :

| | P | $\nu = 10 \times 10 - 1$ |
|-------------------|------|--------------------------|
| χ^2_{α} | 0.01 | 69.23 |
| | 0.5 | 98.33 |
| | 0.99 | 134.64 |

Sedaj, ko imamo napisano tabelo za primerjalne vrednosti, si oglejmo izračunane vrednosti, ki nam jih da posamezen generator:

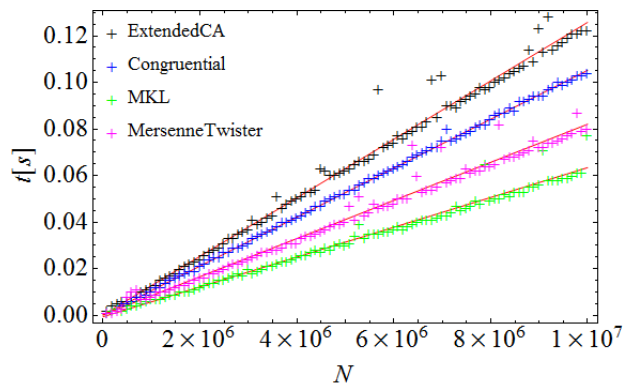
| $N = 100$ | Metoda | Test: 99 ± 14.07 | Metoda | Test: 99 ± 14.07 | $N = 1000$ |
|-----------|------------------|----------------------|------------------|----------------------|------------|
| $S = 10$ | Default | 99.25 ± 14.41 | Default | 98.70 ± 14.63 | $S = 10$ |
| | Kongruenčni | 99.05 ± 13.97 | Kongruenčni | 98.92 ± 14.12 | |
| | Intel MKL | 98.99 ± 14.13 | Intel MKL | 99.18 ± 13.87 | |
| | Mersenne Twister | 99.33 ± 14.39 | Mersenne Twister | 98.65 ± 13.66 | |

| Metoda | Test: 99 ± 14.07 |
|------------------|----------------------|
| Default | 98.39 ± 13.46 |
| Kongruenčni | 98.86 ± 14.12 |
| Intel MKL | 98.84 ± 14.36 |
| Mersenne Twister | 99.04 ± 14.08 |
| $N = 10\,000$ | $S = 10$ |

Tudi v tem primeru rezultati testa niso preveč pretresljivi. Tudi tu zaključimo z rezultatom, da so vsi generatorji kvalitetni, zaključna misel pa je enaka kot pri 1D porazdelitvi, t.j., da razmazanost porazdelitve χ^2 ne pomeni, da je naša metoda slaba, ker je tako ne moremo zavreči ali potrditi z fiksno stopnjo tveganja α , temveč ravno to potrjuje naključnost podatkov.

4 Časovna potratnost

Dodaten parameter, ki tudi kaže kvaliteto generatorja, je njegova časovna poraba, saj si ne želimo takšnih, ki rabijo ogromno časa le za par števil, tudi če so prava naključna števila. Spodnja slika prikazuje časovno porabo testiranih generatorjev v odvisnosti od količine generiranih števil:



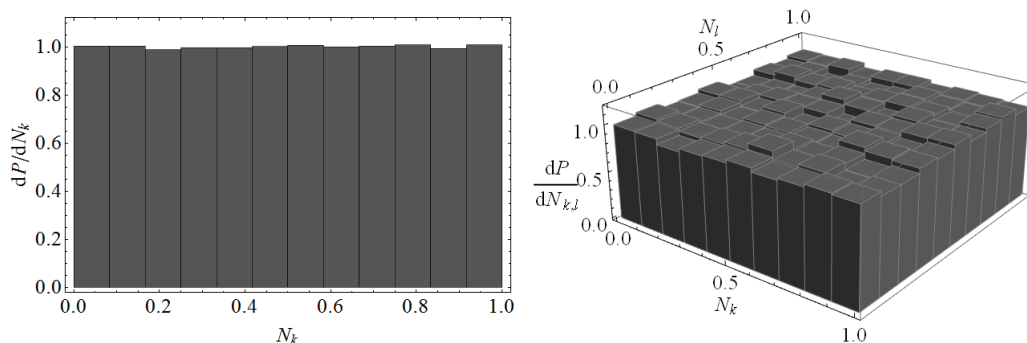
Vidimo, da čas narašča linearno z številom generiranih števil v pri vseh metodah. Za najhitrejšega se je izkazal Intelov MKL generator, ostali pa si potem sledijo po vrsti, kjer je bil najbolj časovno potraten kar privzeti generator. Za vzete količine je MKL generator potreboval dobre 3 s, medtem ko je privzeti generator potreboval dobrih 6, tako da imamo razliko v hitrosti generacije za faktor 2.

5 Slab generator - Randu

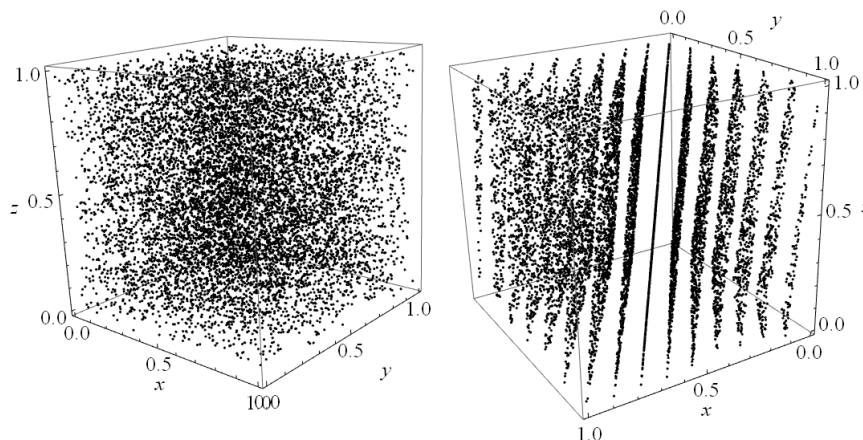
Potreba po naključnih številih je veliko starejša od računalnikov, zato je bilo treba generatorje naključnih števil, po tem ko so se pojavili prvi računalniki, še razviti. Ker se vsak nov izum dlje časa izpopolnjuje, sej je tudi tu zgodilo, da veliko generatorjev ni produciralo najbolj naključno porazdeljena števila. Veliko generatorjev je t.i. kongruenčnih, ki generirajo neko zaporedje oblike

$$x_{i+1} = ax_i + c \pmod{m}. \quad (14)$$

Podjetje IBM je nekoč razvilo generator naključnih števil z imenom **randu**, ki je vseboval parametre $c = 0$, $a = 65539$ in $m = 2^{31}$. Kot so v poročilu že videli, ni dovolj le naključnost števil, ampak morajo biti števila tudi nekolerirana v čimveč možnih dimenzijah, tu pa se je izkazal ta generator kot zelo slab. Spodaj si lahko ogledamo porazdelitve števil v 1, 2 in 3 dimenzijah, kjer sm pri 1D in 2D porazdelitev predstavil z normiranima histogramoma:



Na prvi pogled ni videti korelacij v 1D in v 2D, pa tudi 3D porazdelitev na spodnji sliki je na prvi pogled pravilna, saj enakomerno napolni celotno škatlo, vendar iz drugega zornega kota dobimo precej drugačno sliko:



Vidimo, da so vse naključne trojice naključnih števil na eni od 15 ravnin, ki jih opiše enačba

$$9x - 6y + z = m, \quad m \in [-5, 9]. \quad (15)$$

Za boljšo predstavo ravnin prilagam povezavo do animacije:
<https://dl.dropboxusercontent.com/u/12747812/Animacije/randu.gif>

Del III

Prostorska porazdelitev

1 Naključne smeri v prostoru

Kot naključna števila lahko generiramo tudi kota θ in ϕ , ki sta kordinati v sferičnem koordinatnem sistemu. Takšni pari koordinat predstavljajo koordinato v prostorskem kotu. Ne smemo pozabiti, da moramo pri pretvorbi koordinat paziti na prispevek Jakobijana, kjer velja:

$$\frac{dP}{dx dy dz} = \frac{dP}{dr d\theta d\phi} |\mathcal{J}| = \frac{dP}{dr d\theta d\phi} r^2 \sin \theta. \quad (16)$$

Ker nas zanimajo porazdelitve samo po smeri v prostoru, lahko porazdelitev po radialni koordinati kar odstranimo. Ker je porazdelitev po smeri v prostoru enakomerna, dobimo z upoštevanjem normalizacije identiteto:

$$\frac{dP}{d\Omega} = \frac{d^2 P}{d\phi d(\cos \theta)} = \frac{1}{4\pi}. \quad (17)$$

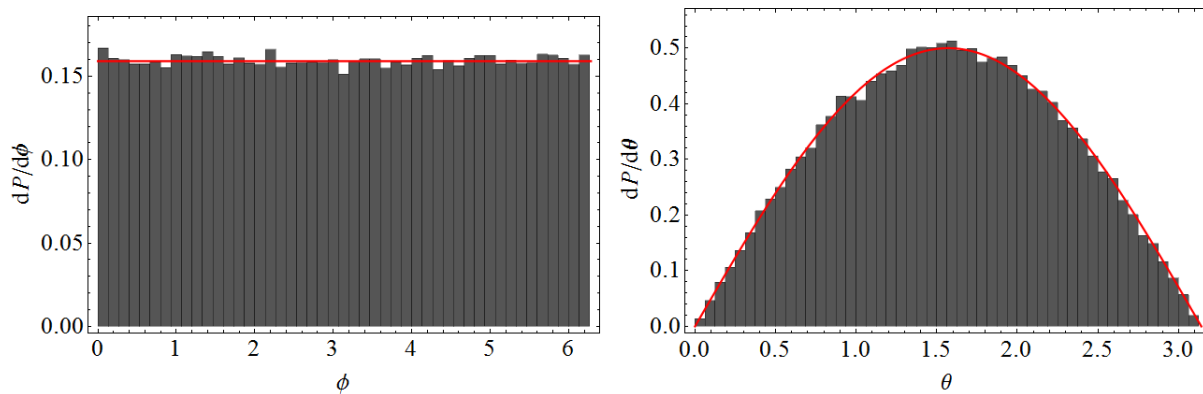
Porazdelitev po spremenljivkah lahko ločimo in tako dobimo način za enakomerno generacijo kotov:

$$\frac{dP}{d\phi} = \frac{1}{2\pi} \rightarrow \frac{dP}{du} = \frac{dP}{d\phi} \left| \frac{d\phi}{du} \right| \rightarrow \phi = 2\pi u, \quad (18)$$

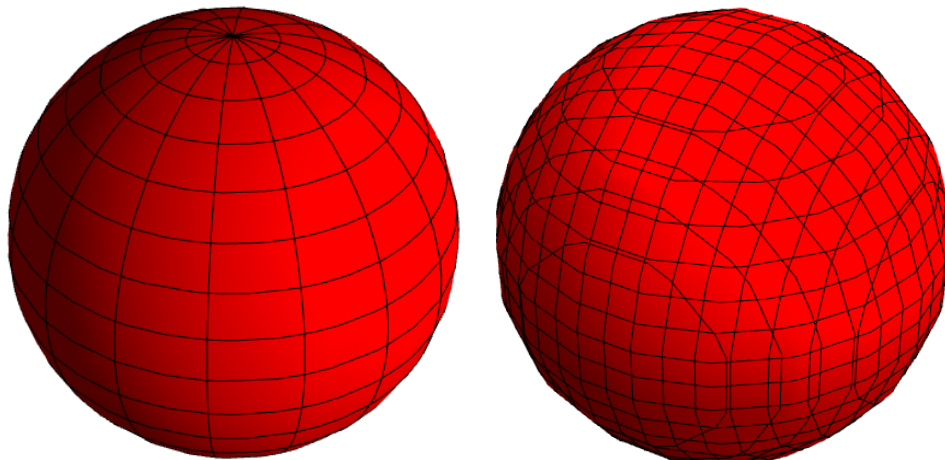
$$\frac{dP}{d(\cos \theta)} = \frac{1}{2} \rightarrow \frac{dP}{dv} = \frac{dP}{d(\cos \theta)} \frac{d(\cos \theta)}{dv} \rightarrow \theta = \cos^{-1}(2v - 1), \quad (19)$$

$$\hookrightarrow \frac{dP}{d\theta} = \frac{1}{2} \sin \theta, \quad (20)$$

kjer sta u in v enakomerno porazdeljeni spremenljivki iz intervala od 0 do 1. S tako definiranimi spremenljivkama u in v sedaj dobimo porazdelitev po kotih ϕ in θ naslednje oblike:



V sferičnih koordinatah s takšno generacijo pokrijemo enotsko površino krogle, kot je prikazano na spodnji sliki z analitično funkcijo in našimi podatki:



Poleg vrednosti povprečnih ϕ in θ si je zanimivo ogledati tudi še kakšne višje momente. Ker so ti avtomatsko vključeni v sferičnih harmonikih, sem pričakovane vrednosti višjih momentov izračunal kar kot:

$$\langle Y_l^m(\theta, \phi) \rangle = \frac{\int_0^\pi \int_0^{2\pi} Y_l^m(\theta, \phi) \sin \theta d\theta d\phi}{\int_0^\pi \int_0^{2\pi} \sin \theta d\theta d\phi} = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} Y_l^m(\theta, \phi) \sin \theta d\theta d\phi,$$

rezultati pa so podani v naslednji tabeli:

| | $\langle \phi \rangle$ | $\langle \theta \rangle$ | $\langle Y_1^0 \rangle$ | $\langle Y_1^1 \rangle$ | $\langle Y_2^0 \rangle$ | $\langle Y_2^1 \rangle$ | $\langle Y_2^2 \rangle$ |
|------------|------------------------|--------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Calculated | 3.13944 | 1.56858 | 0.00027 | -0.00084 | -0.06254 | -0.00037 | -0.00060 |
| Exact | $\pi = 3.14159$ | $\pi/2 = 1.57079$ | 0 | 0 | 0 | 0 | 0 |

Naši rezultati se lepo ujemajo s teoretičnimi vrednostmi. Naša generacija naključnih smeri v prostoru je torej kvalitetna.

2 Sevanje dipola

Pri dipolu ponovno generiramo naključne kore θ in ϕ , vendar za njih v tem primeru velja drugačna porazdelitev, ki je oblike:

$$\frac{dP}{d\Omega} = \frac{d^2P}{d\phi d(\cos \theta)} = C \sin^2 \theta, \quad (21)$$

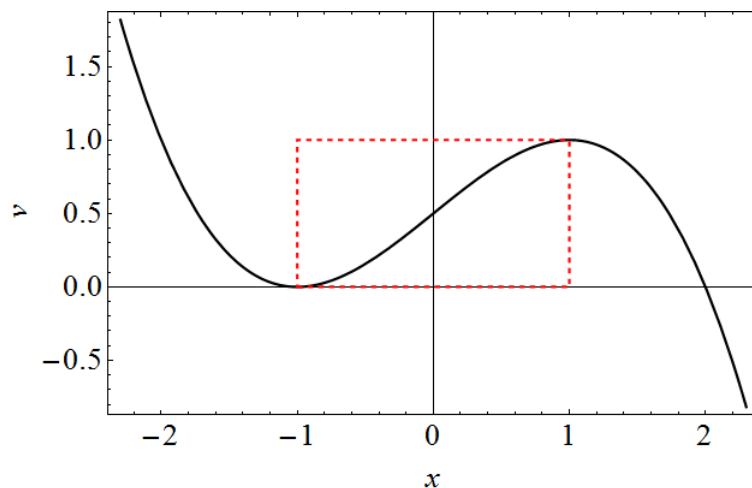
kjer je C normalizacijska konstanta. Pri dipolu torej smeri v prostoru niso več naključno porazdeljene, zato sklepamo, da bo celotni prostorski manjši, porazdelitev pa bo neenakomerna. Normalizacijsko konstanto najdemo, ko izvrednotimo:

$$\int_{\Omega} \frac{dP}{d\Omega} d\Omega = 1 \rightarrow C = \frac{3}{8\pi}. \quad (22)$$

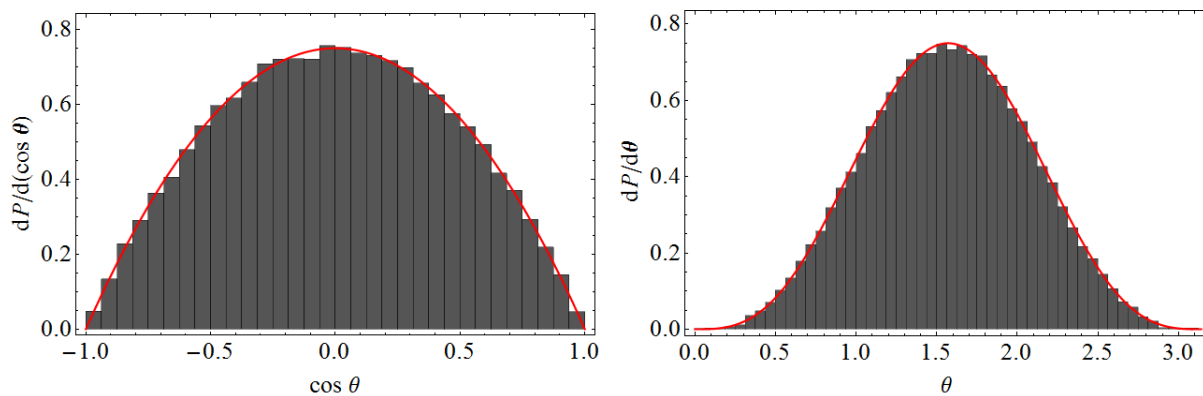
S tem pa določimo prostorski kot v primeru sevanja dipola $\Omega = \frac{8\pi}{3}$. Ker je porazdelitev po kotu ϕ identična tisti iz prejšnjega razdelka, lahko izrazimo porazdelitev po kotu θ kot:

$$\frac{dP}{d(\cos \theta)} = \frac{3}{4} \sin^2 \theta \rightarrow \frac{dP}{dv} = \frac{dP}{d(\cos \theta)} \frac{d(\cos \theta)}{dv} \rightarrow v = \frac{1}{4} (3 \cos \theta - \cos^3 \theta) + D, \quad (23)$$

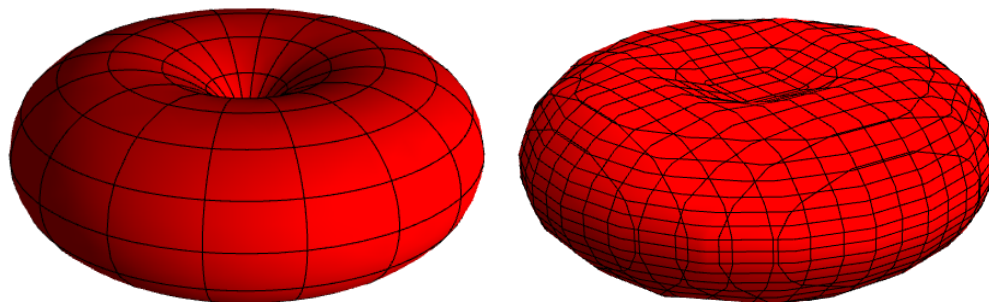
kjer nastavimo $D = 1/2$, tako da je za območje θ med 0 in π naš v v območju med 0 in 1. Vidimo, da ima obliko polinoma:



Polinom je v realnem območju omejen, kot je prikazano na sliki, saj θ nastopa znotraj kosinusa. Vidimo, da takšna funkcija ni enostavno obrnljiva, zato se lahko poslužimo bisekcije, minimizacijske metode, ali pa rešimo problem s kakšnim matematičnim orodjem. Ker je funkcija v relevantnem območju monotono naraščujoča, sem se odločil, da bom izvednotili $\theta(v)$ z numerično minimizacijo, kar se storil tako, da sem na drobno tabeliral vrednosti v , jim priredil rešitev $\theta(v)$, to funkcijo pa nato interpoliral, da sem jo lahko izvednotil tudi v vmesnih točkah. S tako definiranimi spremenljivkama u in v dobimo porazdelitev po kotih ϕ in θ . Porazdelitev po ϕ je enaka kot prej, porazdelitvi po $\cos \theta$ in θ pa imata naslednjo obliko:



Res se porazdelitev po kotu theta spremeni. Čeprav na pogled ne dobimo velikih razlik, so te bolj očitne v sferičnih koordinatah, kjer tokrat z generacijo smeri ne pokrijemo več enotske površine krogle, temveč neko novo torusu podobno obliko, ki je prikazana na spodnji sliki z analitično funkcijo in našimi podatki:



Poleg vrednosti povprečnih ϕ in θ si je tudi tokrat zanimivo ogledati tudi še kakšne višje momente preko sferičnih harmonikov. Pri računanju povprečja z integralom sedaj dobimo zraven nov porazdelitveni faktor:

$$\langle Y_l^m(\theta, \phi) \rangle = \frac{\int_0^\pi \int_0^{2\pi} \frac{3}{4} \sin^2 \theta Y_l^m(\theta, \phi) \sin \theta d\theta d\phi}{\int_0^\pi \int_0^{2\pi} \frac{3}{4} \sin^2 \theta \sin \theta d\theta d\phi} = \frac{3}{8\pi} \int_0^\pi \int_0^{2\pi} \frac{3}{4} \sin^2 \theta Y_l^m(\theta, \phi) \sin \theta d\theta d\phi,$$

rezultati pa so podani v naslednji tabeli:

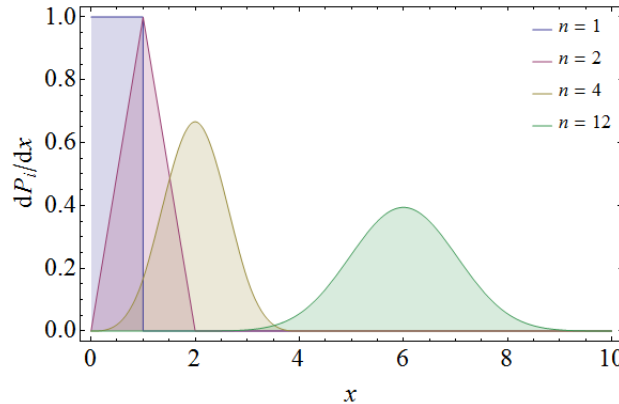
| | $\langle \phi \rangle$ | $\langle \theta \rangle$ | $\langle Y_1^0 \rangle$ | $\langle Y_1^1 \rangle$ | $\langle Y_2^0 \rangle$ | $\langle Y_2^1 \rangle$ | $\langle Y_2^2 \rangle$ |
|------------|------------------------|--------------------------|-------------------------|-------------------------|--------------------------------------|-------------------------|-------------------------|
| Calculated | 3.13911 | 1.56824 | 0.00061 | 0 | -0.139054 | 0.00047 | 0.00009 |
| Exact | $\pi = 3.14159$ | $\pi/2 = 1.57079$ | 0 | 0 | $-\frac{1}{2\sqrt{5}\pi} = -0.12616$ | 0 | 0 |

Tudi tokrat se naši rezultati se lepo ujemajo s teoretičnimi vrednostmi, malce odstopa le Y_2^0 , kar pa bi verjetno popravili z večjo količino generiranih števil. Naša generacija naključnih smeri pri sevanju dipola je torej kvalitetna.

Del IV

Normalna porazdelitev slučajnih števil

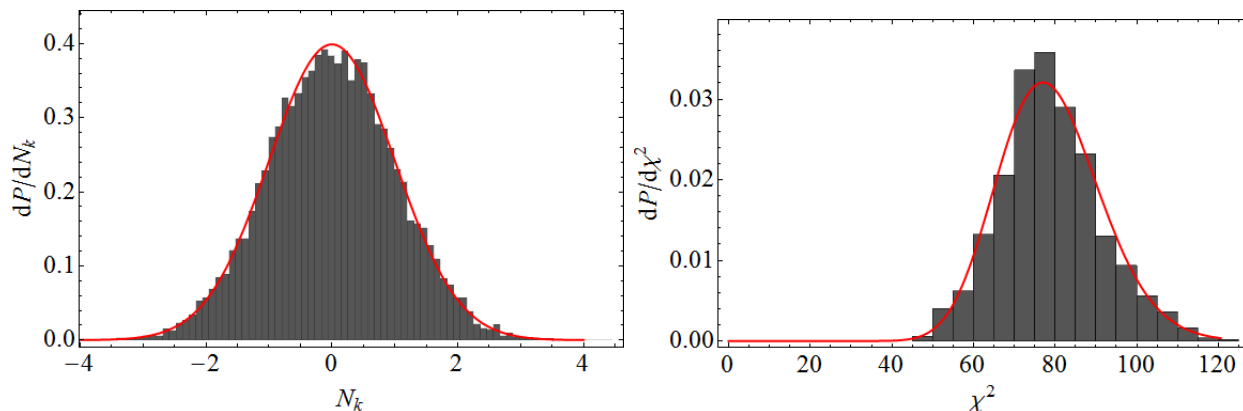
Med generatorji gaussovskih slučajnih števil najdemo Box-Mullerjev in konvolucijski generator. V tem delu smo preizkusili še omenjene generatorje normalno porazdeljenih števil na isti način kot v drugem delu. Najprej nekaj besed o teh dveh metodah generiranja normalno porazdeljenih števil. Box-Mullerjeva metoda je privzeta, zato tu ne bomo izgubljali besed, konvolucijska metoda pa temelji na centralnem limitnem izreku, ki pravi, da je vsota n -tih naključnih porazdelitev konvergira k normalni, ko n limitira proti ∞ , tudi če seštevane vsote niso normalno porazdeljene. Oglejmo si nekaj porazdelitev konvolucijske porazdelitve za nekaj različno dolgih vsot:



Izkaže se, da se gaussovi porazdelitvi še najbolj približa porazdelitev 12 konvolucij, zato v naslednjih izračunih uporabljamo kar to distribucijo.

1 Test χ^2

Tokrat sem teste napravil le za primer $N = 10000$, saj je drugače količina števil premajhna. Pri testu smo števila razporedili v $S = 80$ razredov, nato pa za takšno razdelitev izvednotili χ^2 , kjer smo izračun nato ponovili 1000-krat in ga predstavili v obliki normiranega histograma. Spodnji sliki prikazujeta normirana histograma normalno porazdeljenih števil in vrednosti testov χ^2 :



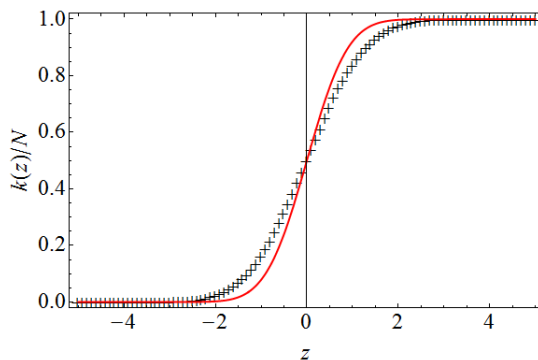
Poleg histogramov sta še teoretični krivulji, katerima se podatki lepo prilegajo. Rezultati testov so prikazani na koncu razdelka.

2 Test K-S

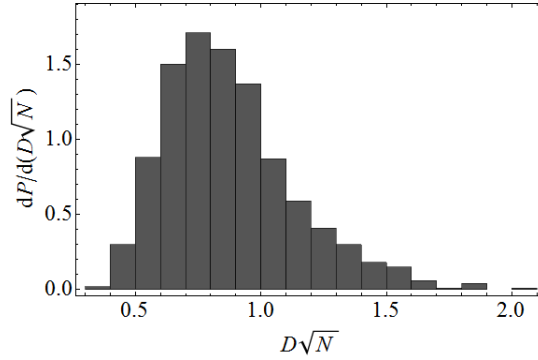
Za test Kolmogorov-Smirnova spet potrebujemo kumulativno porazdelitev naključnih števil. Teoretična kumulativna porazdelitev je v tem primeru kar:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad (24)$$

ki jo moramo še ustrezno premakniti in raztegniti, da se bo skladala z območjem naših podatkov:



Vidimo, da teoretična krivulja malce odstopa od kumulativne porazdelitve podatkov, kar pa bi lahko popravili tako, da bi vzeli večje število podatkov. Iz tega razloga sem tudi v tem primeru vedno računal z $N = 10000$. Za takšno porazdelitev lahko izvednotimo $D\sqrt{N}$ po testu K-S, izračun pa nato ponovimo 1000-krat. Spodaj je prikazan normiran histogram podzdelitve vrednosti $D\sqrt{N}$:



Tudi tu rezultati skačejo po celem verjetnostnem območju, tako da zaključimo z isto mislijo kot v prejšnjih poglavjih.

3 Rezultati testov

Za primerjavo rezultatov ponovno prilagam tabelo relevantnih vrednosti testov χ^2 in K-S za različne stopnje tveganja α :

| | | | | |
|---------------|------|----------------|------------|----------|
| | P | $\nu = 80 - 1$ | | P |
| $\chi^2_{d>}$ | 0.01 | 52.72 | $K-S_{d>}$ | 0.01 |
| | 0.5 | 78.33 | | 0.5 |
| | 0.99 | 111.14 | | 0.99 |
| | | | | 0.441028 |
| | | | | 0.827574 |
| | | | | 1.62762 |

Rezultati testov so sledeči:

| | | |
|---------------|---------------|------------------------|
| $N = 10\ 000$ | Metoda | Test: 79 ± 12.5698 |
| $S = 80$ | Box-Muller | 79.4528 ± 12.8069 |
| | Convolutional | 81.7531 ± 13.0622 |

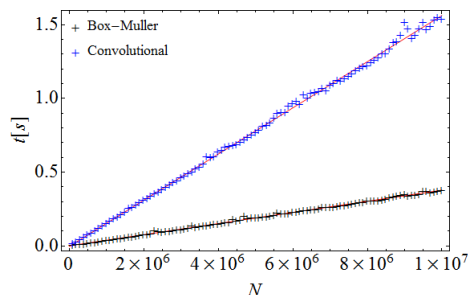
Opazimo, da se iz tega vidika konvolucijski generator ne ujema s teoretično krivuljo kot Box-Mullerjev generator, kar je smiselno, saj porazdelitev ni točno gaussovska. Oglejmo si še rezultate testa K-S:

| | | |
|---------------|---------------|-----------------|
| $N = 10\ 000$ | Metoda | Test K-S |
| | Box-Muller | 0.85 ± 0.25 |
| | Convolutional | 0.94 ± 0.28 |

Tudi ta test nam pokaže, da iz tega vidika konvolucijski generator naključnih števil ni tako dober kot Box-Mullerjev, saj povprečje testa ne leži v okolici 50 %, čeprav pa je še vedno zadovoljiv. Oba generatorja sta uspešno prestala oba testa.

4 Časovna potratnost

Kvaliteto generatorjev z vidika časovne porabe lahko preverimo tudi pri generatorjih normalno porazdeljenih števil, kar prikazuje slika spodaj:



Čeprav imamo tudi tu povsod linearno odvisnost, pa so tu razlike med metodami občutno večje. Ne le da je konvolucijski generator slabši, poleg tega porabi še ogromno dlje časa za generacijo iste količine števil, saj je za vzete količine potreboval kar 80 s, medtem ko je privzeti generator potreboval le 20 s, kar pa je še vedno več od časovne porabe generatorjev uniformno porazdeljenih števil.

Del V

Oddaja časov nalog

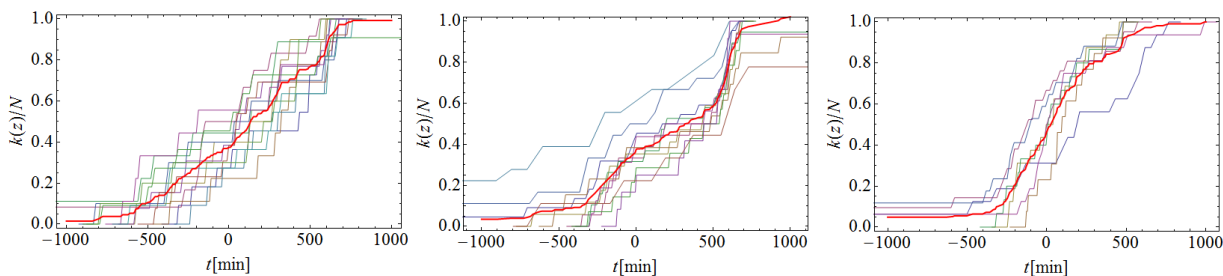
1 Naloga

Dobili smo tudi neobvezno nalogo, kjer preverimo kumulativne porazdelitve časov oddajanja poročil pri predmetu Modelska analiza I. Iz danih datotek sem izluščil podatke in jih nato ločil po letnicah. Najprej sem se želel prepričati, če je so kumulativne porazdelitve oddajnih časov nalog čez leto naključne, ali pa če se časi recimo po nekem trendu daljšajo. Rezultat je bil presenetljiv, saj so bili podatki res porazdeljeni naključno kot si lahko ogledamo na pripeti povezavi do animacije, ki prikazuje porazdelitve v kronološkem zaporedju:

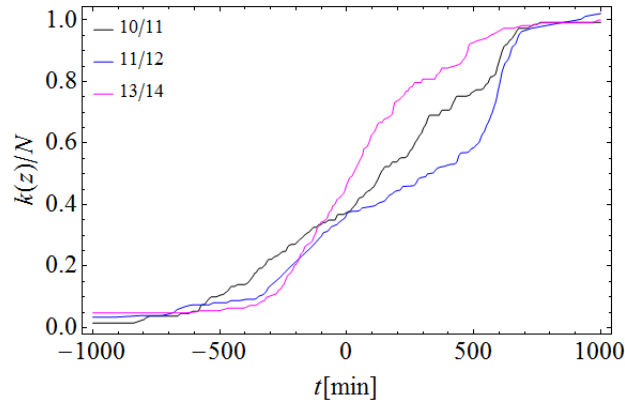
<https://dl.dropboxusercontent.com/u/12747812/Animacije/time.gif>

2 Rezultati

Odločil sem se, da bom skupnim kumulativnim porazdelitvam za posamezno leto priredil neko povprečno porazdelitev, kar sem dosegel tako, da sem interpoliral porazdelitve za posamezno nalogo v določenem letu, interpolacije pa nato seštel. Povprečje, in porazdelitev posameznih kumulativnih porazdelitev si lahko ogledamo na naslednjih treh slikah, ki predstavljajo posamezna šolska leta v kronološkem zaporedju:



Vidimo, da rdeča interpolacijska krivulja dobro opiše povprečje porazdelitev. Sedaj lahko te tri povprečitve primerjamo med sabo in tako sklepamo, če so porazdelitve statistično razločljive:



Ni presenetljivo, da so si vse tri krivulje statistično enake, saj ne vidimo korelacij s šolskim letom. Takšna rešitev je tudi smiselna, saj se pogoji predmeta, fakultete in ostalih dejavnikov, ki bi lahko prispevali, niso spremenili, ali pa se niso spremenili bistveno. Odstopanja lahko pripišemo čisti naključnosti podatkov.

Del VI

Generator rand- π

Zanimivo se mi je zdelo še preizkusiti, kako dober generator naključnih števil je število π , ki ima, kot vemo, neskončno decimalk. Odločil sem se na hitro preveriti, kakšna je 1D, 2D in 3D parov decimalk, kar sem dosegel tako, da sem vzel neceli del števila, sparil skupaj po pet cifer in to deklariral za novo število, nato pa delil z največjim številom v seznamu, tako da sem dobil porazdelitev na območju $[0,1)$. Za konec si le še oglejmo omenjene porazdelitve v 3 različnih dimenzijah, kjer sem general $N = 20000$ števil:

