

MODELSKA ANALIZA I

1999/2000

2. naloga - Slučajna števila po zahtevani porazdelitvi

Andrej Studen, MA-FI

20. oktober 1999

1 Žrebanje smeri v prostoru

Smer v prostoru navadno opišemo s kotoma θ in φ . Tako je smiselno iskanje preliti v sferične koordinate, tam pa imamo zaradi transformacije:

$$p(x_1, x_2, x_3)dx_1dx_2dx_3 = p(y_1, y_2, y_3)\mathcal{J}(T)dx_1dx_2dx_3$$

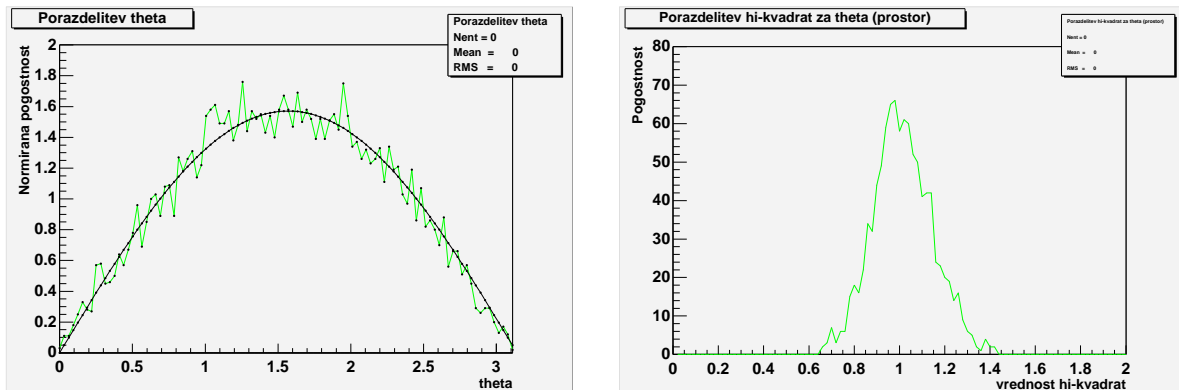
spremenjeno porazdelitev. Če je bila prej porazdelitev enakomerna po prostoru, bomo v sferičnih koordinatah dobili še dodaten člen zaradi $\mathcal{J}(T)$, Jacobijeve determinante transformacije. Seveda je ta za prehod v kote enaka $\sin(\vartheta)$, tako da lahko zapišemo:

$$\frac{dw}{d\Omega} = C$$
$$\frac{d^2w}{d\varphi d\vartheta} = C \sin \vartheta,$$

kjer je C normirna konstanta. To slednje lahko prepišemo tudi v:

$$\frac{dw}{d\varphi} \frac{dw}{d \cos \vartheta} = C,$$

saj smo motečo kotno funkcijo pospravili v diferencial. Vemo, da je $C = \frac{1}{4\pi}$, tako da smo že dobil recept.



Slika 1: Porazdelitev po ϑ ni več enakomerna, pač pa se zaradi prehoda v drugačne koordinate razsuje. Seveda je sorazmerna $\sin \vartheta$, kar nam kaže tudi leva slikica, ko okoli prave vrednosti skače naša porazdelitev. Desno pa vidimo, da lahko tudi za take porazdelitve izračunamo χ^2 in da je porazdelitev presenetljivo dobra.

Izberemo kot φ med 0 in 2π z verjetnostjo $\frac{1}{2\pi}$, kar nam oklesti konstanto na $\frac{1}{2}$, nato pa še na intervalu od -1 do 1 izberemo kosinus ϑ s konstantno verjetnostjo $\frac{1}{2}$ zaradi razširitve intervala, pa smo opravili.

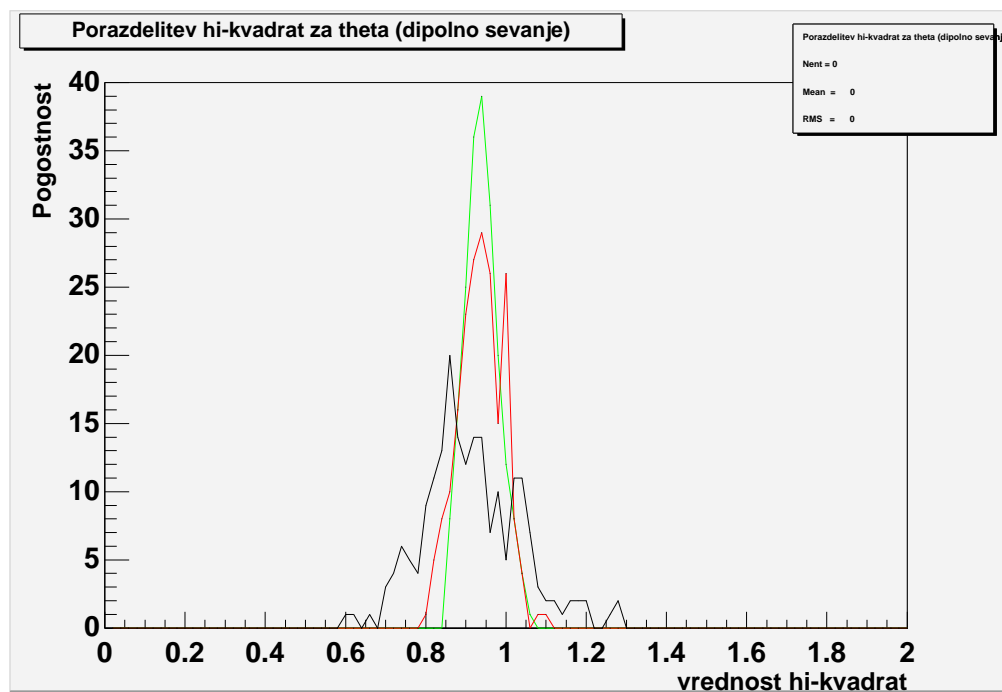
Tabela 1: Primerjava pravih momentov in njihovih odmikov za generiranje naključne smeri v prostoru

moment	prava	povprečje	prava σ	σ
fi	3.14159	3.13117	3.28987	3.23971
theta	1.5708	1.56867	0.467401	0.463123
y10	0	0.0017314	0.333333	0.330726
y20	-0.25	-0.00781255	0.8	0.789767
y11	0	0.0115674	0.333333	0.336961
y21	0	0.00490468	0.0666667	0.067839
y22	0	0.00275678	0.266667	0.269491

Seveda interval izbire naključnega števila brez težav raztegnemo (množimo), s tem tudi zmanjšamo verjetnost na konstantnem intervalu, in premaknemo, s čimer verjetnostne porazdelitve na samem intervalu ne spreminjamo. Tako je:

$$\varphi = 2\pi\xi_1 \quad \text{in} \quad \vartheta = \text{ArcCos}(2\xi_2 - 1)$$

Opazoval sem nekatere momente pri tej porazdelitvi. Ob že klasičnih $\langle \varphi \rangle$ in $\langle \vartheta \rangle$ še momente krogelnih funkcij $\langle Y_{lm} \rangle$. Rezultate kaže preglednica. Velika neujemanja so predvsem pri σ . Mogoče sem to narobe izračunal ali pa je 10000 števil premalo za dobro oceno odmika. Povprečja so namreč kar prava.

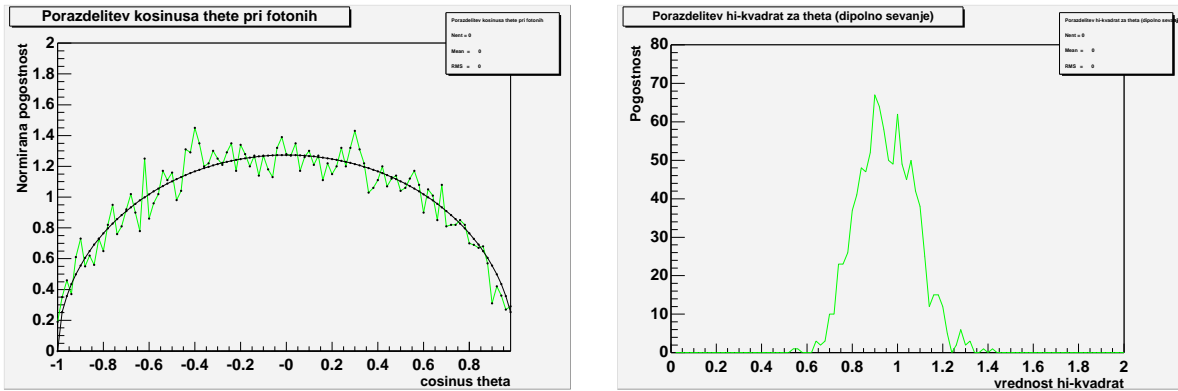


Slika 2: Porazdelitev χ^2 za enakomerno porazdelitev po φ . Tri krivulje na sliki so bile narisane za različne vrednosti N (števila poskusov za vsak χ) in M (števila predalčkov), pri stalnem razmerju $\frac{N}{M} = 100$. Najvišja in najožja krivulja je tista z največjim N (10^6), je pa ta tudi hkrati najbolj leva. Zdi se, da gre povprečje krivulje k vrednostim, manjšim od 1 za velike N . Ožanje je posledica teorije, ki napove širino porazdelitve χ^2 kot nekaj, sorazmernega inverznemu korenu M . To se na sliki lepo vidi. V splošnem se zdi porazdelitv po kotu φ odlična.

Tabela 2: Primerjava izračunanih in dejanskih momentov nekaterih znanih funkcij za porazdelitev, sorazmerno $\sin \vartheta$. Ujemanja so tako v prvem kot v drugem moemntu, torej gre za smiselne porazdelitve.

moment	prava	povprecje	prava σ^2	σ^2
fi	3.14159	3.13117	3.28987	3.23971
theta	1.5708	1.57264	0.322467	0.319641
y10	0	-0.00151301	0.25	0.247787
y20	-0.25	-0.256633	0.5625	0.556749
y11	0	0.0115482	0.375	0.379108
y21	0	-0.00476163	0.0625	0.0632648
y22	0	0.00176223	0.3125	0.315372

2 Porazdelitev fotonov dipolnega sevanja po smereh v prostoru



Slika 3: če zahtevamo, da je porazdelitev po prostoru sorazmerna $\sin \vartheta$, se porazdelitev po ϑ tudi močno spremeni. Tako je na levi sliki prikazana porazdelitev po *kosinusu* ϑ , ki gre kot $\sin(\text{ArcCos}(\cos \vartheta))$, kar je nekaj v zlahti z elipso. Na desni pa je prikazana porazdelitev po χ^2 , tokrat za porazdelitev po ϑ . Ta je sorazmerna $\sin^2 \vartheta$, kar pa se na sliki za χ^2 seveda ne vidi. Vendar je porazdelitev kljub temu prav lepa.

Zdaj preprosto zapišemo:

$$\frac{dw}{d\Omega} = \sin \vartheta,$$

torej bo porazdelitev po φ spet konstantna na intervalu $[0, 2\pi)$, medtem ko bo za porazdelitev po ϑ veljalo:

$$\frac{dw}{d\vartheta} = \frac{2}{\pi} \sin^2 \vartheta.$$

Ta izraz je celo integrabilen, dobimo grdo klobaso:

$$w(\vartheta) = \frac{1}{2}\vartheta - \frac{1}{4} \sin 2\vartheta,$$

ki pa seveda ni obrnljiva. Vsaj ne analitično. Ker pa si želimo simulatorja dipolnega sevanja, ko bomo obrnili numerično. Vsaj jaz sem jo kar z bisekcijo. Integral je namreč monotona funkcija in ima zato rešitev za vsako število med 0 in 1. Namesto $w(\vartheta)$ pišemo naključno število ξ nekje na intervalu zaloge vrednosti intervala; za normirano funkcijo kot našo je to pač vrednost med 0 in 1, nato pa z bisekcijo poiščemo pripadajočo ϑ in imamo smer. Primerjave orazdelitev dajejo slike, momenti pa so navedeni v preglednici. Težave so podobne kot pri uniformni prostorski porazdelitvi.

3 Generiranje normalne (Gaussove) porazdelitve

Obstajata dva najpreprostejša načina za določitev naključnih števil, porazdeljenih po Gaussu. Eden je seštevanje, torej:

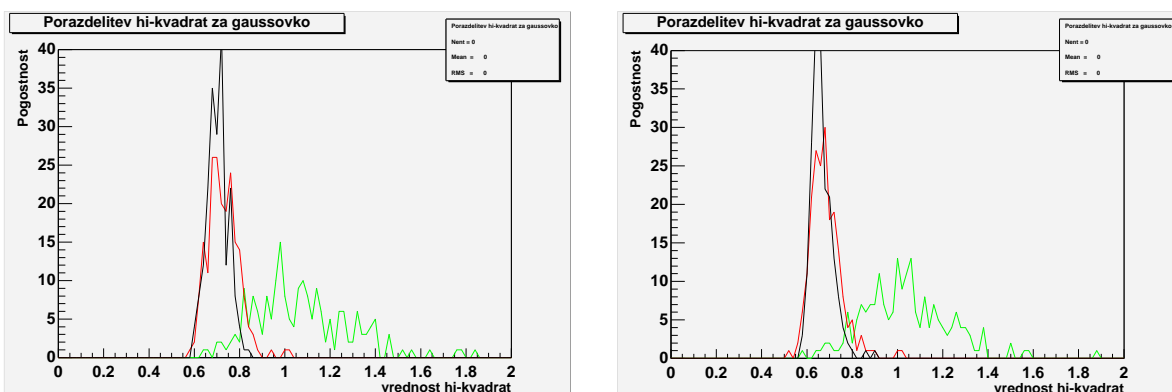
$$\gamma = \sum_{i=1}^N (-1)^i \xi_i,$$

kjer (-1) skrbijo, da se nam povprečje ne premakne, število seštetih naključnih števil pa je odvisno od željene σ . Ker je za enakomerno porazdelitev $\sigma = \frac{1}{\sqrt{12}}$, je za σ_G potrebno sešteti 12 števil. Seveda to pomeni, da pod -6 in nad 6 ni zadetkov, a je tam verjetnost tako majhna, da se to porazdelitvi komajda pozna. Drug način pa izhaja iz integrala. V eni dimenziji je gaussovo funkcijo težko integrirati, saj pridemo do numerične *erf*, ki je zato težko obrnljiva. Vendar pa lahko dvodimenzionalno gaussovko s spremembo koordinat privedemo do napotka:

$$\gamma_1 = \sqrt{-2 \ln \xi_1} \cos(2\pi \xi_2)$$

$$\gamma_2 = \sqrt{-2 \ln \xi_1} \sin(2\pi \xi_2)$$

Brez težav se da pokazati, da sta γ_1 in γ_2 neodvisni naključni števili in da za obe velja neodvisno, da sta porazdeljeni po Gaussu. Primerjalni testi obeh načinov so na slikah. Očitno je, da je drugi hitrejši in natančnejši, prvi pa služi le kot trening.



Slika 4: K porazdelitvi χ^2 prispeva tudi način računanja t vrednosti. V njej nastopa v vsakem členu vsote v imenovalcu vrednost σ^2 . To lahko vzamemo iz izmerjenih vrednosti, ali pa iz teoretičnih podatkov, obakrat je $\sigma^2 = \sqrt{N_i}$, $N_i = h * N * f(x_i)$, h je širina predalčka, N število poskusov, $f(x)$ pa prava porazdelitev. Razliko obeh načinov kažeta zgornji sliki. Levo so to koreni iz števila zadetkov, desno funkcijske vrednosti. Na obeh slikah pa so prikazane porazdelitve za različne N in M (glej podnapise pri prejnjih slikah), $N/M=100$, $N=(10^5, 5 \cdot 10^5, 10^6)$. Krivulje so vse ožje, povprečja pa se zaskrbljujoče manjšajo. Vzrok je mogoče to, da zaradi manipulacije z enakomernimi porazdelitvami, ki naj bi nas pripeljale blizu gaussovkam, preveč omejimo razmah na^{ih} naključnih števil in so krivulje preveč pravilne.

Pri obeh pa je zanimivo, da se nam χ^2 seli proti 0 za velike vrednosti N(števila poskusov) in M(števila predalčkov) ob stalnem razmerju $\frac{N}{M}$. Zdi se, da postajajo funkcije zaradi težkih predpisov preveč pravilne in izgubljammo naključne napake. A kdo bi vedel. LEpo je videti tudi ožanje krivulje, skladno s teorijo, za naraščajoče M.

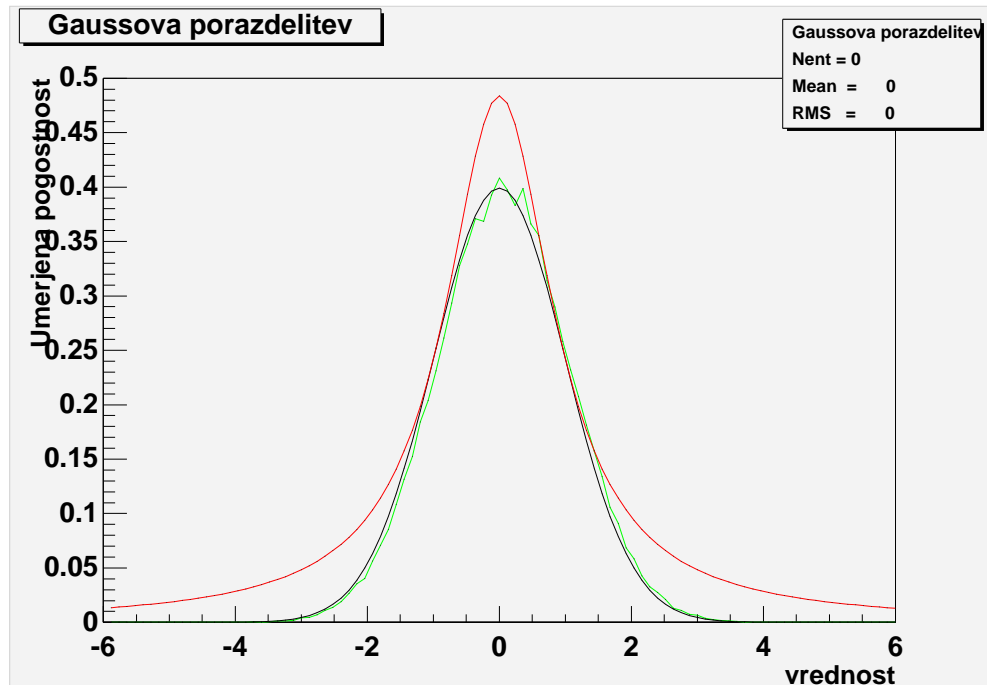
4 Gaussova porazdelitev z zavračanjem

Različne porazdelitve (binomsko, Poissonovo) lahko generiramo z zavračanjem. Za to rabimo samo krivuljo, ki na vsem definicijskem območju prekriva našo funkcijo, ki je integrabilne in katere integral je (preprosto)

obrnjiv. Za Gaussovo je dobra krovna funkcija Lorentzova, katere parametre lahko določimo tako, da bo prekrila gaussovo. Tako recimo Lorentz oblike:

$$\sqrt{\frac{2}{e\pi}}(1+x^2)^{-1}$$

dobro prekrije Gausso za $\bar{x} = 0$ in $\sigma = 1$.



Slika 5: Pri generiranju z zavračanjem izbiramo vrednosti, ki so porazdeljene po zgornji (Lorentzovi) krivulji, zavračamo pa tiste, pri katerih drugo naključno število (y) pade izven območja, ki ga omejuje spodnja krivulja. Tako dobimo porazdelitev na sliki, ovito okrog spodnje prave krivulje. Kot se vidi tudi iz razmerja ploščin veliko število zavrnemo (približno 5 za 10 dobrih)

Prave naključne vrednosti pa generiramo takole:

- izberemo vrednost med 0 in A, kjer je A vrednost integrala krovne funkcije po celem prostoru
- z inverzom integrala poiščemo vrednost x , pri kateri je integral dosegel našo izbrano vrednost
- izberemo še drugo naključno vrednost y med 0 in vrednostjo krovne funkcije v x . Tako izbran par (x,y) je porazdeljen enakomerno med krovno funkcijo in absciso.
- če leži y tudi v prostoru med absciso in našo iskano funkcijo, obdržim x , sicer ga odvržem. Pravzaprav je y neke vrste senzor. Za dani x je verjetnost, da ga bomo vzeli, ravno enako razmerju med vrednostma krovne in iskane funkcije. Z naključnim y pa preverimo, če naj x vzamemo. Tudi pri razmerjih blizu 1 se nam lahko zgodi, da kak x zavrnemo, a le malokrat. Tako je res verjeti, da so izžrebani x porazdeljeni po iskani krivulji.

Zanimivo je, da z zavračanjem ne izgubimo tako veliko časa. Z omenjeno funkcijo porabimo za vsako naključno število celo manj časa kot za seštevanje 12 števil. Tako pride sicer okoren način do dobre interpretacije. Razmerja med rezultati podaja tabela, primerajavo χ^2 pa slike. Vidimo, da je zavračanje malo manj natančno kot ostala dva načina, a še vedno sprejemljivo.

Tabela 3: Primerjava povprečnih časov, potrebnih za določitev enega naključnega števila z normalno porazdelitvijo. Izstopa `gasdev.c`, ki računa direktno in je zato najhitrejša rutina, medtem ko sta ostali dve podobno počasni. Mogoče je treba dati `gauss1.c`, ki računa s seštevanjem, nekaj rezerve, ker je napisana za splošen σ in \bar{x} , kar zahteva nekaj if stavkov v vsakem koraku. Presenetljivo hiter pa je `gasrej.c`, čeprav zavrne več kot tretjino vrednosti.

rutina	čas [μ s]
<code>gauss1.c</code>	5.48
<code>gasdev.c</code>	0.80
<code>gasrej.c</code>	4.24

5 Naključne točke po volumnu valja

Spet nam je napoti nesrečen Jacobian. Kot že vemo, je:

$$p(x_1, x_2, x_3)dx_1dx_2dx_3 = p(y_1, y_2, y_3)\mathcal{J}(T)dx_1dx_2dx_3,$$

kar pomeni, da bo porazdelitev, enakomerna v kartezičnem sistemu (kakršne so vse, ki so enakomerne po volumnu) v drugem sistemu dobila še kakšen faktor. Tako bomo mi dobili:

$$\frac{dw}{dV}(r, \varphi, \vartheta) = \frac{3}{\pi h R^2} r,$$

pa še prostor za te spremenljivke ni še pravokotnik. A bližje smo. Najprej izločimo φ , ki se giblje v svojem prostoru. Seveda je porazdelitev enakomerna na $[0, 2\pi)$, tako ob noramlizaciji oklestimo našo konstanto. Ostane nam:

$$\frac{dw}{dS}(r, z) = \frac{6}{hR^2} r,$$

kjer je S površina med osema r in z. Nesemo r v diferencial na levi strani in zamenjamo koordinate ($r \rightarrow r^2$), da dobimo:

$$\frac{dw}{dS}(r^2, z) = \frac{3}{hR^2},$$

saj smo 2 porabili za diferencial $d(r^2) = 2rdr$, S pa zdaj pomeni ploskev med z in r^2 . Pridobili smo konstantno porazdelitev po ploskvi, ki pa jo omejuje parabola meje stožca. Če bi bili v valju, bi lahko izbrali katerikoli par (r^2, z) po enakomerni verjetnosti v vsaki smeri. Tako pa imamo omejujočo krivuljo. Linearna funkcija, ki povezuje r in z:

$$\frac{r}{R} + \frac{z}{h} = 1,$$

se spremeni v parabolo:

$$r^2 = u = R^2 \left(1 - \frac{z}{h}\right)^2$$

oziroma:

$$z = h \left(1 - \frac{\sqrt{u}}{R}\right).$$

Tu je takoj očitno, da obstajata dve poti izbire naključnega elementa naše površine. Lahko smatramo, da je z funkcija nad u, ali pa da je u funkcija nad z. Pri obeh primerih ravnamo enako. Določimo integral funkcije $u(z)$ ali $z(u)$, poiščemo inverz, izberemo vrednost med 0 in vrednostjo integrala po celem območju, z inverzom najdemo ustrezno vrednost neodvisne spremenljivke, nato pa s pomočjo funkcije določimo meje izbiranja druge številke po enakomerni porazdelitvi. Ker gre ravno za diferenciale ploščine, bo izbrana verjetnost že normirana na volumen stožca. Če vzamemo $z(u)$, dobimo nekoliko zoprn integral:

$$w(u) = \frac{3u}{R^2} - \frac{2}{R^3} u\sqrt{u} = \xi_1,$$

Tabela 4: Primerjava časov, potrebnih za določitev ene naključne točke v stožcu. Pri `point1.c` se delamo, da je podana krivulja $u(z)$, pri `point2.c` pa, da je podana krivulja $z(u)$ (tu uporabljam bisekcijo). Razlika je očitna.

rutina	čas{ μ s}
<code>point1.c</code>	3.0
<code>point2.c</code>	21.2

ki ga obrnemo z bisekcijo, z pa dobimo z drugim naključnim številom:

$$z = \xi_2 \left(-\frac{h}{R} \sqrt{u_\xi} + h \right) \quad \text{seveda je} \quad r = \sqrt{u_\xi},$$

kjer je u_ξ z bisekcijo dobljeni inverz integrala pri naključnem številu ξ . Druga pot ($u=u(z)$) se izkaže za bolj obetavno, najdemo celo enostaven inverz integrala. Ta je namreč enak:

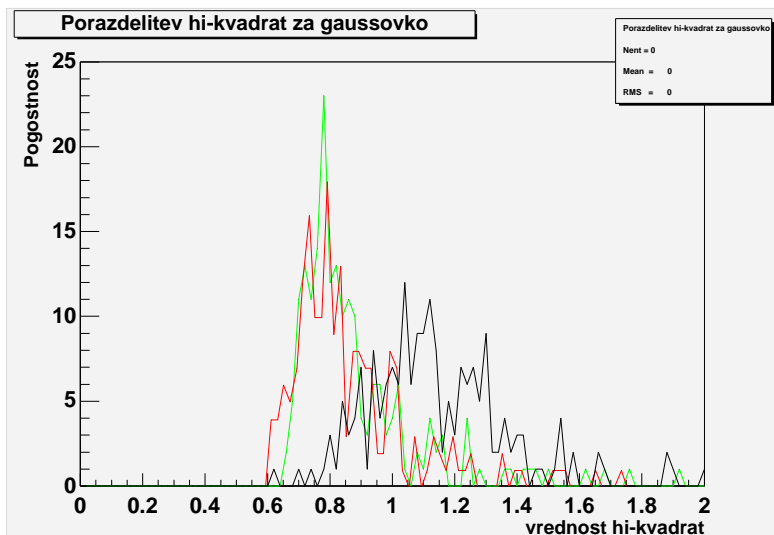
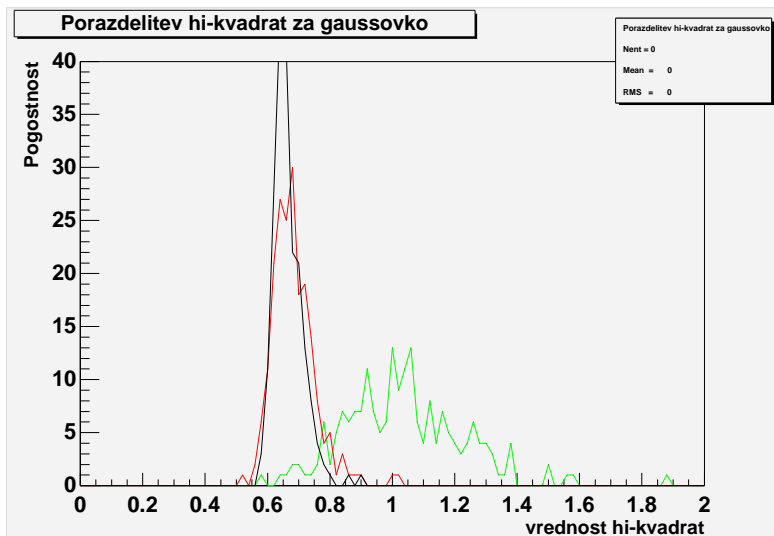
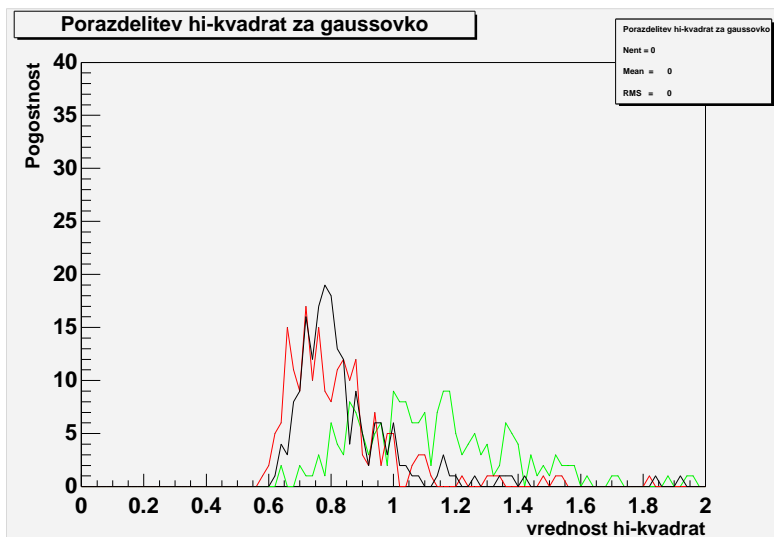
$$w(z) = \left\{ \left(1 - \frac{z}{h} \right)^3 - 1 \right\},$$

kar nam da za r in z :

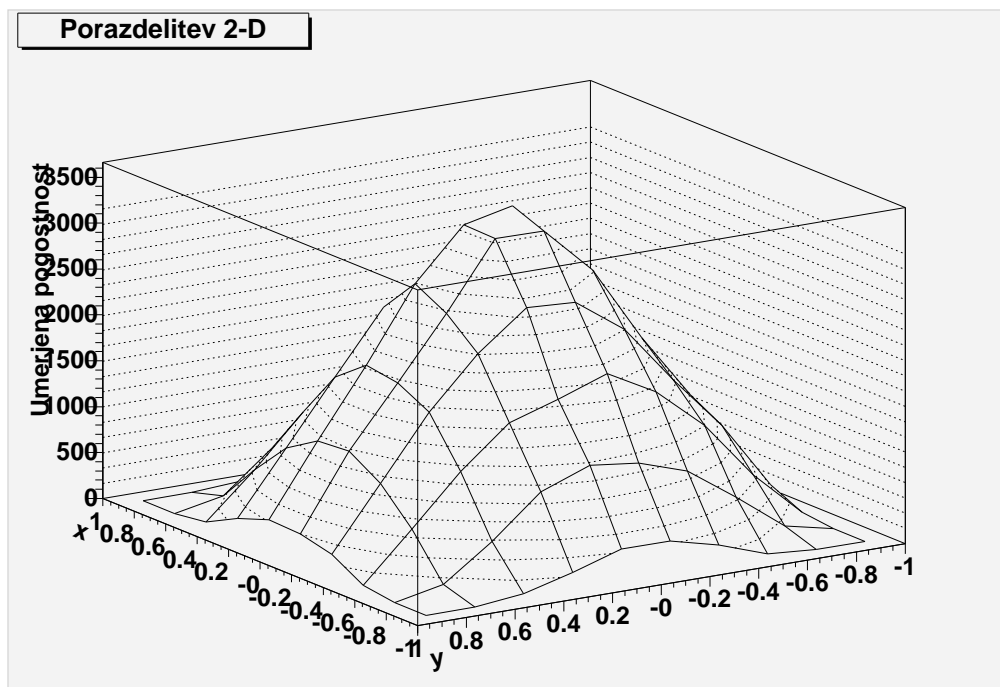
$$z_\xi = h \left(1 - \sqrt[3]{1 - \xi_1} \right)$$

$$r = R \sqrt{\xi_2} \left(1 - \frac{z_\xi}{h} \right).$$

Kot vidimo, smo morali r vedno še koreniti. Tako dobimo po ploskvi res enakomerno porazdelitev, hkrati pa upoštevamo še tretjo dimenzijo. Slednji postopek je končen, za vsako število porabi natančno tri korake, vsakega za izračun svojega števila. Deluje neodvisno od razmerja $\frac{h}{R}$ in najhitreje, kar lahko, če mora vrniti tri številke. Zato se mi zdi, da vprašanje v zvezi z razmerjem ni več v igri. Če bi oba načina zahtevala bisekcijo, bi mogoče z razvojem v katerem od primerov bisekcijo močno omilili. Tako pa ne vidim razloga, da bi navedeno razmerje kakorkoli vplivalo na računanje po drugi poti in ga navajam kot optimalnega za izračun točk v stožcu.



Slika 6: Primerjava porazdelitev po χ^2 za tri naše načine določanja števil po Gaussovi funkciji. To dopolnjuje tabelo o izmerjenih časih. NA vsaki sliki so tri krivulje, ki ustrezajo trem različnim številom poskusov (N) pri določanju χ^2 ($N=10^5, 5 \cdot 10^5, 10^6$), razmerje med številom predalčkov (M) in N pa sem obdržal konstantno. Vidi se padajoči trend vseh treh krivulj, vzrok je mogoče v pretirani natančnosti ali pa nenatančnem integriranju pri primerjavi vrednosti v χ^2 testu. Po vrsti zgoraj formula, sredi seštevanje in spodaj zavračanje.



Slika 7: če gledamo projekcijo izžrebane točke na ravnino xy , bo realtivna pogostnost teh točk ravno opisala stožec nad ravnino, če prav izbiram točke. Ker res dobim stožec, je to pravilna metoda. Stožec je videti dokaj pravilen. ($N=10000$, $m(\text{delilne točke na stranico})=10$)