Líkindaaðferðir

Reikniverkefni 2 og 3
Kristinn Gylfason

18. mars 2000

Reikniverkefni 2

Tilgangur verkefnisins er að kanna helstu markgildissetningu tölfræðinnar. Búa á til 200 sýni af

$$Z = \frac{X_0 + X_1 + ... + X_{N-1}}{\sqrt{N}}$$

þar sem X_i er jafndreifð slembibreyta með væntigildi núll og staðalfrávik einn fyrir $N = 5$, $N = 10$ og $N = 25$. Því næst á að finna meðaltal og fervik Z, teikna dreifirit og bera saman við þéttifall samsvarandi normaldreifingar fyrir hvert tilvik.

Lausn

Til þess að hægt sé að bera dreifiritin saman við líkindaþéttifall þarf að norma dreifiritin. Það er gert með því að deila fjölda staka í hverju bili með heildarfjölda staka sinnum breidd bilsins, eins og er útskýrt á bls 73 í Cooper og McGillem.

Matlab-kóðinn fyrir verkefnið er sýndur í forriti 1. Myndir 1, 2 og 3 sýna svo normuðu dreifiritin sem beðið er um ásamt Gaussferli. Inn á myndirnar eru einnig merkt tilsvarandi meðaltöl, fervik Z vigursins og kí-kvaðrat gildi (sem beðið er um í næsta lið). Meðaltal og fervik er reiknað fyrir eitt tiltekið sett af tölvugefnum slembigildum en kí-kvaðrat gildin eru meðaltal kí-kvaðrat gilda fyrir 100 mismunandi sett af slembigildum.

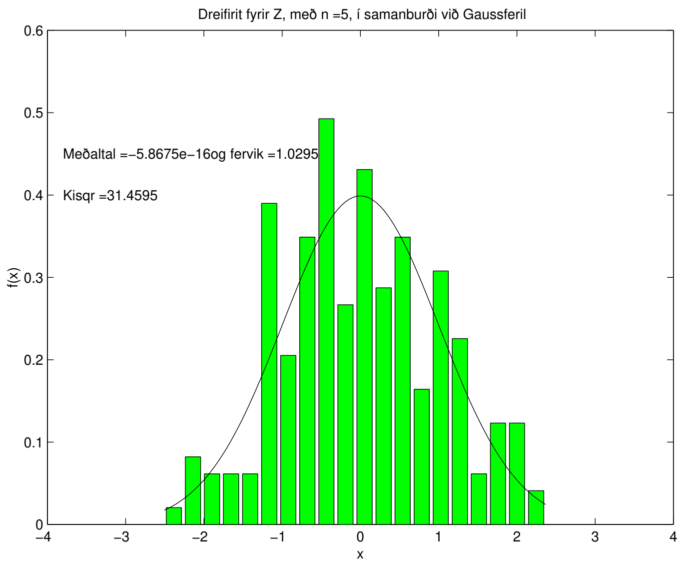
Helsta markgildissetning tölfræðinnar segir að summa af mörgum óháðum slembibreytum með sama líkindaþéttifall, væntigildi og fervik hafi Gaussískt líkindaþéttifall með sama væntigildi og fervik og upphaflegu breyturnar. Þessi er enda raunin með þau dæmi sem eru tekin hér. Reyndar virðist það ekki bæta niðurstöðuna að hækka n. Það skýrist þó frekar með skekkju í kí-kvaðrat prófinu.

Forrit 0.1 Matlab-kóði fyrir reikniverkefni 2 og kí-kvaðrat próf á Z.

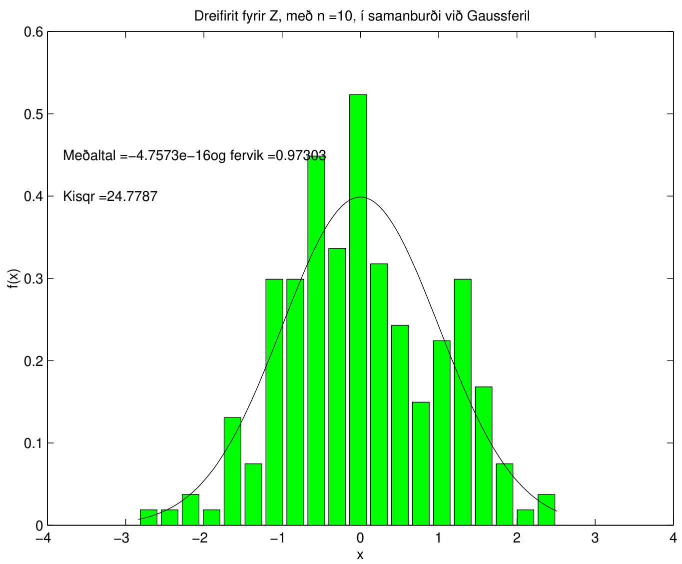
```
%reikniverkefni 2 og 3 í líkindaaðferðum

clear all;

%global constants
elements = 200;
bins = 20;
tests = 100;
n = [5 10 25];

for j=1:length(n)  %loop over n

 kisum = 0;

 for m=1:tests     %loop to get a ki-square average

   z = zeros(elements,1);
   x = rand(elements,n(j));

   %add random variables and scale outcome
   for i=1:n(j)
      x([1:elements],i) = ( x([1:elements],i) - mean(x([1:elements],i)) ) / std(x([1:elements],i));
      z = z + x([1:elements],i);
   end;
   z = z/sqrt(n(j));

   %normalise histogram
   [distr,center] = hist(z,bins);      %seperate z into bins
   with = ( max(z)-min(z) ) / (bins-1);    %find with of bins
   binprob = distr/(elements*with);    %find probability of each bin
   xscale = linspace(min(z),max(z));   %linear vector in the z space
   gausscurve = gaussian(xscale,0,1);

   kisqr = 0;
   for k=1:bins
      term = ( distr(k) - elements*with*gaussian(center(k),0,1) ).^2 / (elements*with*gaussian(center(k),0,1));
      kisqr = kisqr + term;
   end;

   kisum = kisum + kisqr;

 end;

 kisum = kisum / tests;

 %plot graph
 figure(j);
 bar(center,binprob, 'g');
 title(strcat( 'Dreifirit fyrir Z, með n = ', num2str(n(j)), ', í samanburði við Gaussferil'));
 text(-3.8, 0.45, strcat( 'Meðaltal =  ' ,num2str(mean(z)), 'og fervik =  ', num2str(var(z)) ) );
 text(-3.8, 0.4, strcat( 'Kisqr =  ' ,num2str(kisum) ) );
 xlabel('x');
 ylabel('f(x)');
 axis([-4 4 0 0.6]);
 hold on
 plot(xscale, gausscurve,'k');
 hold off

end;


fnames = ['fig1' 'fig2' 'fig3'];
for h=1:length(fnames)
   print( '-deps', '-f', fnames(h) )
end
```

Reikniverkefni 3

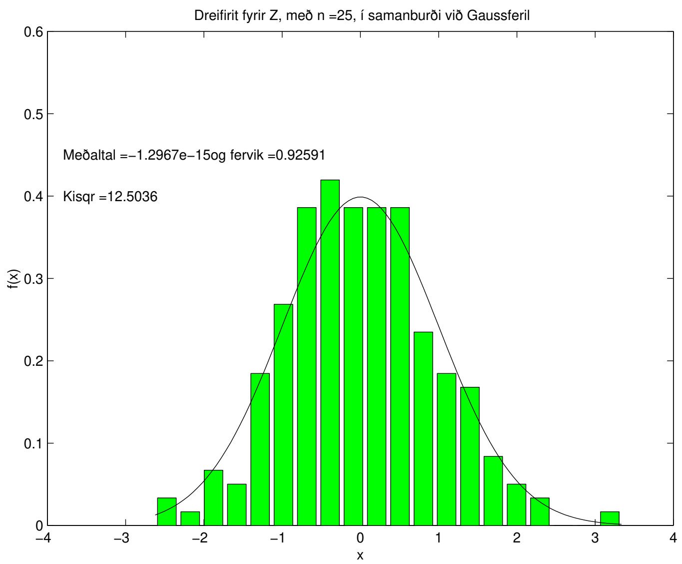
Tilgangur verkefnisins er að skoða kí-kvaðrat próf. Fyrst á að nota randn() fallið til að búa til normaldreifðan vigur af slembigildum með staðalfrávik tvo og


Dreifirit fyrir Z, með n =5, í samanburði við Gaussferil
Meðaltal =−5.8675e−16og fervik =1.0295
Kisqr =31.4595
f(x)
x

Mynd 1:


Dreifirit fyrir Z, með n =10, í samanburði við Gaussferil
Meðaltal =−4.7573e−16og fervik =0.97303
Kisqr =24.7787
f(x)
x

Mynd 2:



Mynd 3:

væntigildi þrjá og prófa svo með kí-kvaðrat prófi hvort vigurinn nálgi Gussferil með sama fervik og meðaltal. Að lokum á að prófa öll þrjú tilvikinn í reikniverkefni 2 með kí-kvaðrat prófi.

Lausn

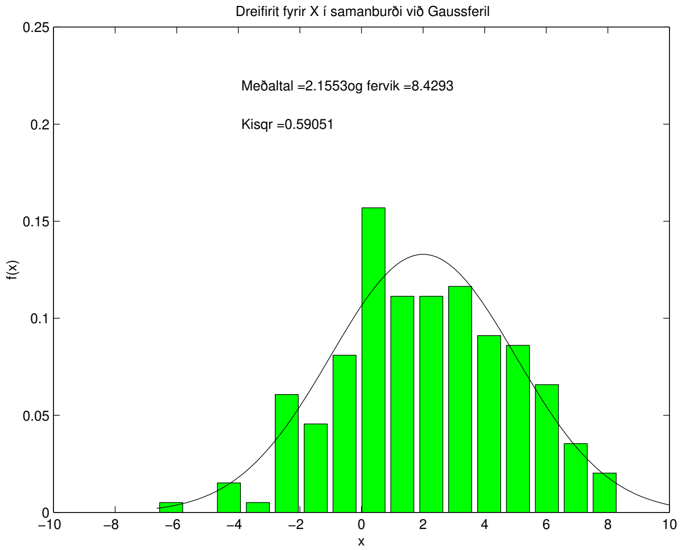
Forrit 0.2 Matlab-kóði fyrir reikniverkefni 3.

```
%reikniverkefni 3 í líkindaaðferðum

clear all;

%global constants
elements = 200;
bins = 20;
tests = 100;

kisum = 0;

for i=1:tests

   x = 3*randn(1,elements) + 2;

   %prepare histogram
   [distr,center] = hist(x,bins);      %seperate z into bins
   with = ( max(x)-min(x) ) / (bins-1);    %find with of bins
   binprob = distr/(elements*with);    %find probability of each bin
   xscale = linspace(min(x),max(x));   %linear vector in the z space
   gausscurve = gaussian(xscale,2,3);

   kisqr = 0;
   for k=1:bins
      term = ( distr(k) - elements*with*gaussian(center(k),3,2) ).^2 / elements*with*gaussian(center(k),3,2);
      kisqr = kisqr + term;
   end;

   kisum = kisum + kisqr;

end;

kisum = kisum / tests;

%plot graph
figure(1);
bar(center,binprob, 'g');
title('Dreifirit fyrir X í samanburði við Gaussferil');
text(-3.9, 0.22, strcat( 'Meðaltal =  ' ,num2str(mean(x)), 'og fervik =  ', num2str(var(x)) ) );
text(-3.9, 0.20, strcat( 'Kisqr =  ' ,num2str(kisum) ) );
xlabel('x');
ylabel('f(x)');
axis([-10 10 0 0.25]);
hold on
plot(xscale, gausscurve,'k');
hold off

print -deps fig4 -f1
```

Matlab kóðinn fyrir verkefnið er sýndur í forriti 2 og á mynd 4 sést drefirit fyrir vigurinn í samanburði við Gaussferil. Inn á myndina er merkt kí-kvaðrat gildið fyrir prófið. Þar sem $k = 20$ og við höfum tvær skorður (á væntigildi og staðalfráviki) eru svigrúmsvíddirnar $v = 20 - 2 = 18$. Tafla gefur því

$$\chi^2_{0,05;18} = 28,87$$

Þar sem χ^2_0 er minna en 28,87 er tilgátan staðfest. Fyrir tilvikin í reikniverkefni 2 fæst sama niðurstaða.


Dreifirit fyrir X í samanburði við Gaussferil
Meðaltal =2.1553og fervik =8.4293
Kisqr =0.59051
f(x)
x
0.25
0.2
0.15
0.1
0.05
0
−10
−8
−6
−4
−2
0
2
4
6
8
10

Mynd 4: