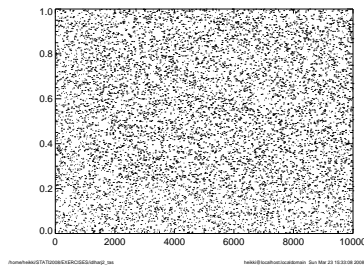


TILASTOLLISET MENETELMÄT TÄHTITIEESSÄ

IDL-harjoitus 2: Esimerkkivastaukset, Heikki Salo 27.3.2008

1. Luo N luvun otos välille $[0, 1]$ tasaisesti jakaantuneita satunnaismuuttujia X (merkitään $X_i \sim \text{TAS}(0, 1)$) käyttäen IDL:n `randomu`-funktioita.

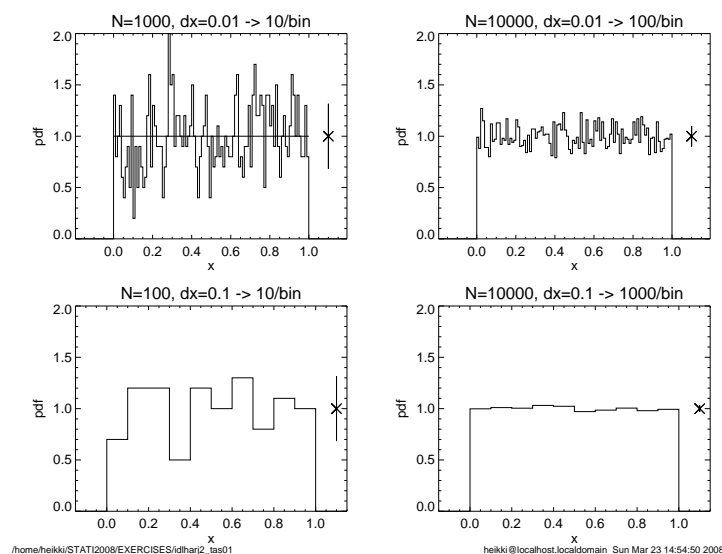
Kokeile esim. IDL> `plot, randomu(seed, 10000), psym=3`



Tämä antaa plotin jossa y-akselina on satunnaisluvun arvo (x-akselilla on luodun satunnaisluvun järjestysnumero).

- Plottaa `histo_f`-proseduurin avulla lukujen jakauma, jakamalla väli $[0, 1]$ esim. 10, 100, 1000,... väliin. Piirrä samaan kuvaan tasaisen jakauman teoreettinen tiheysfunktio. Kokeile eri N :n arvoja.

Ohjelma `idlharj2_tas01.pro` lue eo. jakauman mukaisia satunnaislukuja ja plottaa luotujen satunnaislukujen jakauman histogrammina. Ohessa esimerkkejä eri N :n ja jakovälien dx arvoilla. Karkea arvio satunnaisuuden aiheuttamalle suhteelliselle vaihtelulle on $\pm 1 / \sqrt{N_{bin}}$, jossa N_{bin} ilmoittaa kuinka montaa satunnaislukua histogrammin yksi pylväs keskimäärin vastaa. (arvio seuraa siitä että lukumäärä on Poisson-jakaantunut, keskiarvo N_{bin} , hajonta $\sqrt{N_{bin}}$; arvio merkitty kuviin pystyviivalla).



Huom: `histo_f` normeeaa pylväiden¹ yhteenlasketun pinta-alan yksikköpinta-alaan, eli se approksimoi tiheysfunktioita (edellyttäen että annetut rajat kattavat koko jakauman (voidaan ohittaa /noscale-keywordilla \Rightarrow plottaa lukumäärät kullakin jakovälillä ilman normeerausta).

```

;-----
;idharj2_tas01
;-----
;yksinkertainen plottaus
program='idharj2_tas'
ps=-1
psdirect,program,ps
plot,randomu(seed,10000),psym=3
psdirect,program,ps,/stop
;-----
program='idharj2_tas01'
ps=-1
psdirect,program,ps

;tasainen jakauma valilla 0-1
;valitaan eri N = satunnaislukujen lukumaara ja plotataan histogrammi.
;histo_f laskee eri osavaleille osuvien arvojen lukumaarat
;ja normeeraa ne niin etta histogrammin pinta-ala = 1
;-> approksimoi tiheysfunktia
;Jotta normeeraus oikein, jakauman on rajoituttava annettujen rajojen valiin
;histogram from x1 to x2 with step dx
;xx,yy return calculated values
;histo_f,x,x1,x2,dx,xx,yy,

nwin
!p.multi=[0,2,2]
!y.range=[0,2]
!y.style=1
!p.charsize=0.7
;-----
N=10001
dx=0.01
x=randomu(seed,n)
title='N=1000, dx=0.01 -> 10/bin'
histo_f,x,-.2,1.2,dx,xx,yy
plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title=title,xs=1
;plotataan samaan kuvaan arvio yksittaisten pylvaiden vaihteluista:
;
; DELTA(NBIN)      sqrt(NBIN)      1
; ----- = ----- = -----
;      NBIN          NBIN          sqrt(NBIN)
;
xarg=[1,1]*1.1
yarg=xarg*0.+1
nbin=n*dx
dyarg=xarg*0.+sqrt(1./nbin)
oploterr,xarg,yarg,dyarg

;teoreettinen tiheysjakauma
xteo=findgen(101)/100.
yteo=xteo*0.+1
oplot,xteo,yteo,lines=0
;-----
N=100001
dx=0.01
x=randomu(seed,n)
title='N=10000, dx=0.01 -> 100/bin'
histo_f,x,-.2, 1.2,dx,xx,yy
plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title=title,xs=1

nbin=n*dx
dyarg=xarg*0.+sqrt(1./nbin)

```

```

oploterr,xarg,yarg,dyarg
;-----
N=1001
dx=0.1
x=randomu(seed,n)
title='N=100, dx=0.1 -> 10/bin'
histo_f,x,-.2, 1.2,dx,xx,yy
plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title=title,xs=1

nbin=n*dx
dyarg=xarg*0.+sqrt(1./nbin)
oploterr,xarg,yarg,dyarg
;-----
N=100001
dx=0.1
x=randomu(seed,n)
title='N=10000, dx=0.1 -> 1000/bin'
histo_f,x,-.2,1.2,dx,xx,yy
plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title=title,xs=1

nbin=n*dx
dyarg=xarg*0.+sqrt(1./nbin)
oploterr,xarg,yarg,dyarg

!p.multi=0
psdirect,program,ps,/stop
end

```

- Miten luot satunnaislukuja välillä $[a, b]$, mikä on luomasi otoksen keskiarvo ja keskihajonta? Miten nämä suhtautuvat tasaisen jakauman teoreettiseen keskiarvoon ($\mu = \frac{b+a}{2}$) ja keskihajontaan ($\sigma = \frac{b-a}{\sqrt{12}}$). Vaikuttaako N otoksen keskihajontaan?

Välille $[a, b]$ tasaisesti jakautuneita lukuja saadaan muunnoksella

$$y = (b - a)x + a$$

jossa x on tasaisesti jakaantunut välille $[0, 1]$.

```

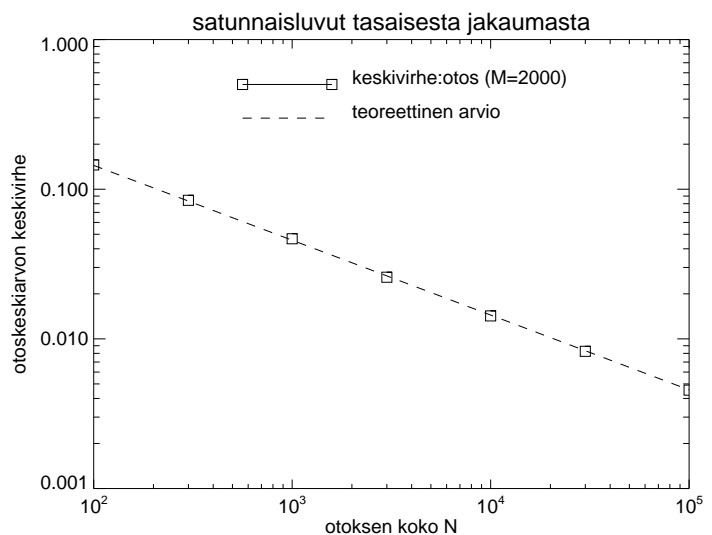
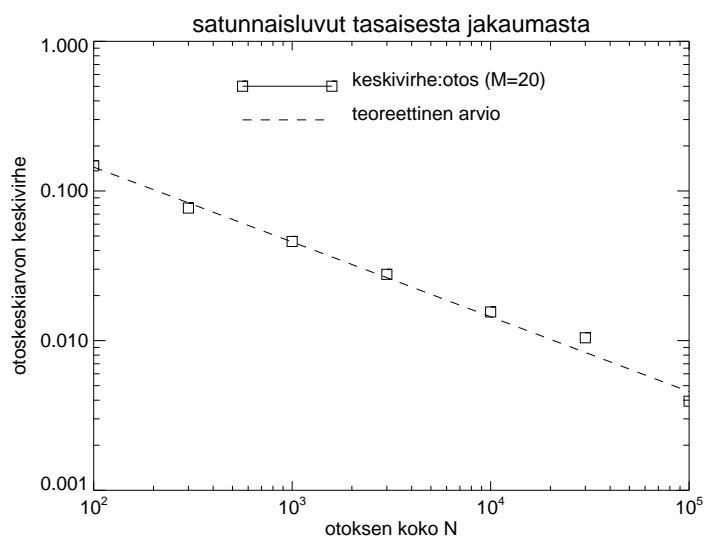
IDL> a=10
IDL> b=15
IDL> print,mean((b-a)*randomu(seed,100)+a),(b+a)/2.
      12.4469      12.5000
IDL> print,mean((b-a)*randomu(seed,100000)+a),(b+a)/2.
      12.5032      12.5000
IDL> print,stdev((b-a)*randomu(seed,100)+a),(b-a)/sqrt(12.)
      1.46380      1.44338
IDL> print,stdev((b-a)*randomu(seed,100000)+a),(b-a)/sqrt(12.)
      1.44160      1.44338

```

Otoskeskiarvo ja otoshajonta eivät sinänsä riipu N :stä (niiden odotusarvo sama riippumatta N :n arvosta), mutta satunnainen vaihtelu pienenee N :n kasvaessa.

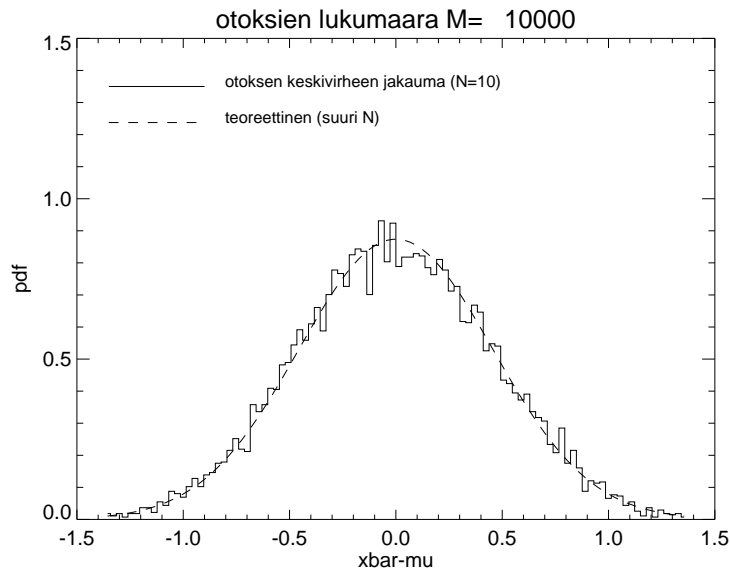
- Totea käyttämällä eri N :n arvoja, että otoksen keskiarvo \bar{X} lähestyy tasaisen jakauman teoreettista keskiarvoa μ otoksen N :n kasvaessa. Arvioi poikkeaman $\bar{X} - \mu$ keskihajontaa (= ”keskiarvon keskivirhe”) luomalla M kappaletta N :n luvun otoksia, ja laskemalla saaduista \bar{X}_j , $j = 1, \dots, M$ niiden otoshajonta σ_μ . Miten σ_μ suhtautuu teoreettiseen arvioon σ / \sqrt{N} , jossa σ on tasaisen jakauman keskihajonta.

(Otoskeskiarvojen ja otoshajontojen laskemiseen IDL:ssä on `mean` ja `stdev` funktiot. Suositus on käyttää uudempaa `moment`-funktioita, joka palauttaa keskiarvon, varianssin, vinokkuuden ja huipukkuuden)



/home/heikki/STAT2008/EXERCISES/IDLharj2_tas_otos

heikki@canopus.fysik.yo.oulu.fi Thu Mar 27 10:41:17 2008



```

;-----
;idharj2_tas_otos.pro
;-----
ps=0
program='idharj2_tas_otos'
psdirect,program,ps

;Tutkitaan tasaisesta jakaumasta otettujen satunnaislukujen keskiarvon
;suppenemista kohti teoreettista arvoa otoskoon kasvaessa.
;Tasaisen jakauman teoreettinen keskiarvo=mu, hajonta=sigma

;Valitaan eri otoskokoja N_tab=[N1, N2,...],
;lasketaan otoskeskiarvo xbar
;Eroituksen xbar-\mu hajontaa kutsutaan keskiarvon keskivirheeksi

;Teoreettinen arvio = sigma/sqrt(N)
;Tutkitaan 'kokeellisesti': luodaan M kappaletta kutakin N luvun otosta
;ja taltioidaan xbar-mu kullekin M, ja lopuksi nain taltioitujen
;erotusten otoshajonta taulokkoon otossigma_tab

;valitaan mielivaltaiset jakauman ala- ja ylaraja
a=5. & b=10.
mu=0.5*(a+b) & sigma=(b-a)/sqrt(12.)

;otoskoot
N_tab=[1001,3001,10001,30001,100001,300001, 1000001]
;otoshajonnat tallennetaan taulukkoon
otossigma_tab=n_tab*0.
;kutakin otosta luodaan M kappaletta
M=20

for in=0,n_elements(n_tab)-1 do begin
  N=n_tab(in)
  xbar_tab=fltarr(m)
  for im=0,m-1 do begin
    x=randomu(seed,n)*(b-a)+a
    xbar_tab(im)=mean(x)-mu
  endfor
endfor

```

```

        otossigma_tab(in)=stdev(xbar_tab)
    endfor

    nwin
    plot,n_tab,otossigma_tab,/ylog,/xlog,$
        title='satunnaisluvut tasaisesta jakaumasta',$
        xtitle='otoksen koko N',ytitle='otoskeskiarvon keskivirhe',psym=6
    oplot,n_tab,sigma/sqrt(n_tab),lines=2
    stn=(M='+strtrim(string(m),2)+'')
    label_data,0.25,0.9,[' keskivirhe:otos '+stm,' teoreettinen arvio'],$
        psym=[6,0],lines=[0,2]
    psdirect,program,ps,/stop

;-----
;havainnollistetaan vielä keskiarvon keskivirheen merkitystä
;verrataan xbar-mu jakaumaa
;ja gaussista jakaumaa jonka hajonta sigma/sqrt(N)

    program='idharj2_tas_otos2'
    psdirect,program,ps

    M=10000          ; monta otosta, jotta niistä saataisiin hyvä jakauma
    N=10             ; otoskoolla ei valia (totea)

    xbar=fltarr(m)
    for im=0,m-1 do begin
        x=randomu(seed,n)*(b-a)+a
        xbar(im)=mean(x)-mu
    endfor
    otossigma=stdev(xbar_tab)

    x1=-3*sigma/sqrt(n)
    x2= 3*sigma/sqrt(n)
    dx=(x2-x1)/100.
    histo_f,xbar,x1,x2,dx,xx,yy

    nwin
    yr=[0,max(yy)*1.5]
    plot,xx,yy,psym=10,xtitle='xbar-mu',ytitle='pdf',yr=yr,title='otoksien lukumaara M='+string(M)

;approksimoidaan gaussisella jakaumalla jonka
;keskiarvo mu_gauss=E(xbar-mu)=0
;hajonta sigma_gauss=sigma/sqrt(N)

    sigma_gauss=sigma/sqrt(N)
    oplot,xx,1./sqrt(2.*!pi*sigma_gauss^2)*exp(-0.5*(xx/sigma_gauss)^2),lines=2

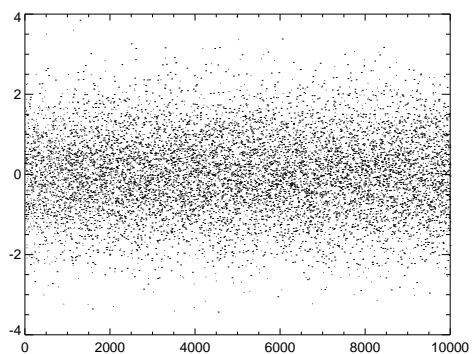
    stn=(N='+strtrim(string(n),2)+'')
    label_data,0.05,0.9,[' otoksen keskivirheen jakauma '+stn,' teoreettinen (suuri N)'],$
        psym=[0,0],lines=[0,2],size=.8

    psdirect,program,ps,/stop
end

```

Luo N kappaleen otos normaalijakaumaa noudattavia satunnaismuuttujia ($X_i \sim N(0, 1)$) käyttäen IDL:n `randomn`-funktiota.

```
IDL> plot,randomn(seed,10000),psym=3
```



Tarkista otoksen keskiarvo ja hajonta. Kuinka luot normaalijakauman, jonka keskiarvo on μ ja keskihajonta σ ?

Lukuja saadaan muunnoksella

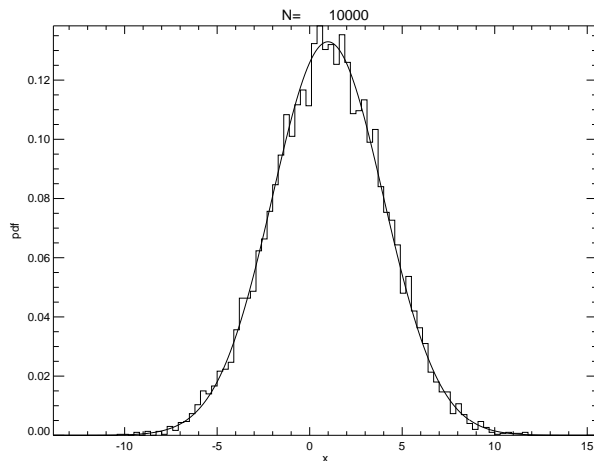
$$y = \sigma x + \mu$$

jossa x on normeerattua normaalijakaumaa noudattava muuttuja.

```
IDL> sigma=3
IDL> mu=1
IDL> print,mean(sigma*randomn(seed,10000)+mu)
0.946455
IDL> print,stdev(sigma*randomn(seed,10000)+mu)
2.98254
```

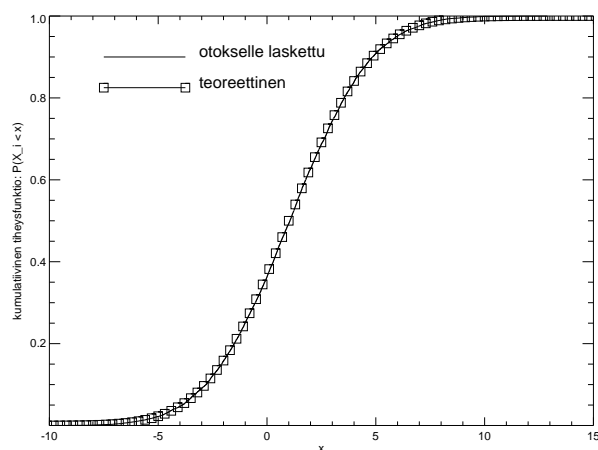
-Plottaa saamasi jakauma eri $N:n$ arvoilla, yhdessä teoreettisen Gaussisen tiheysfunktion kanssa.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]$$



-Plottaa otoksesi kumulatiivinen tiheysjakauma. (Vihje: `sort`-funktio antaa taulukon alkioiden suuruusjärjestyksen, eli `x(sort(x))` järjestää $x:n$ alkioit suuruusjärjestykseen; Kokeile mitä saat kun kirjoitat “`plot,x(sort(x)), lindgen(n)/(float(n)-1)`”, jossa “ $n=n_elements(x)$ ” on otoksen alkioiden lukumäärä)

Vertaa Gaussisen jakauman teoreettiseen kumulatiiviseen tiheysfunktioon $\Phi(x)$ (Muista `gauss_pdf`-funktio.)




```

;-----
;idlhaj2_gauss
;-----
;Gaussisesti jakautuneet satunnaisluvut
;keskiarvo mu, hajonta sigma
;-----

;yksinkertainen plottaus
;normeerattu normaalijakauma: mu=0, sigma=1
    program='idlhaj2_gauss'
    ps=0
    psdirect,program,ps
    plot,randomn(seed,10000),psym=3
    psdirect,program,ps,/stop

;-----
;valitaan eri N = satunnaislukujen lukumaara ja plotataan
;histogrammi.+ verrataan teoreettiseen jakaumaan
    program='idlhaj2_gauss1'
    ps=0
    psdirect,program,ps
    nwin
    !p.charsize=0.7

    sigma=3.
    mu=1.
    N=100001
    title='N='+string(N)
    x=sigma*randomn(seed,n)+mu

;histogrammi valilla -5*sigma -> 5*sigma
    x1=mu-5*sigma
    x2=mu+5*sigma
    dx=(x2-x1)/100.
    histo_f,x,x1,x2,dx,xx,yy
    plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title=title,xs=1

;teoreettinen tiheysjakauma

    xteo=x1+(x2-x1)*findgen(1001)/1000.
    yteo=1./sqrt(2.*!pi)/sigma*exp(-0.5*(xteo-mu)^2/sigma^2)
    oplot,xteo,yteo,lines=0

    psdirect,program,ps,/stop

;-----
;kumulatiivinen tiheysfunktio

    sigma=3.
    mu=1.
    N=100001
    x=sigma*randomn(seed,n)+mu

    program='idlhaj2_gauss2'
    ps=0
    psdirect,program,ps
    nwin
; plotataan otoksen luvut suurrujarjestyksessa, vastaavaksi y:n
; arvoksi laitetaan 0/n,

    plot,x(sort(x)), (lindgen(n)+1)/float(n),xtitle='x',$
        ytitle='kumulatiivinen tiheysfunktio: P(X_i < x)',col=2

```

```

;teoreettinen, kayttaen IDL gauss_pdf-funktiota
xteo=findgen(101)/100.*10.-5.
yteo=gauss_pdf(xteo)

oplot,xteo*sigma+mu,yteo,col=3,psym=6

label_data,0.1,0.9,['otokselle laskettu','teoreettinen'],col=[2,3],psym=[0,6],lines=[0,0]
psdirect,program,ps,/stop
end

```

3. Normaalijakaantuneita satunnaismuuttujia voidaan luoda tasaisen jakauman avulla seuraavasti: Valitaan

$$x_1 \sim \text{TAS}(0, 1) \text{ ja}$$

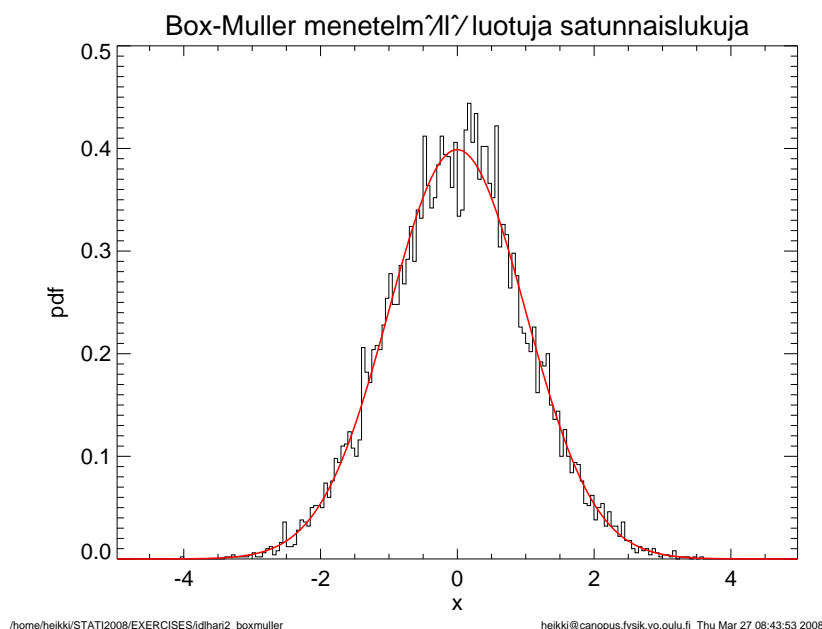
$$x_2 \sim \text{TAS}(0, 1)$$

ja muodostetaan luvut

$$y_1 = \sqrt{-2 \log x_1} \cos 2\pi X_2 \text{ ja}$$

$$y_2 = \sqrt{-2 \log x_1} \sin 2\pi X_2$$

Totea että menetelmä toimii (myös IDL:n `randomn`-funktio käyttää tätä Box-Muller menetelmää).



```

;-----
;idlharj2_boxmuller
;-----
;Gaussisesti jakautuneiden satunnaislukujen luonti Box-Muller menetelmalla
;-----
    program='idlharj2_boxmuller'
    ps=0
    psdirect,program,ps,/color
;tasaisesti jakaantuneet
    n=10000
    x1=randomu(seed,n)
    x2=randomu(seed,n)

;nama ovat normaali jakauman (mu=0, sigma=1)
;mukaisesti jakaantuneet
    y1=sqrt(-2*log(x1))*cos(2*pi*x2)
    y2=sqrt(-2*log(x1))*sin(2*pi*x2)

;y1 histogrammi valilla -5*sigma -> 5*sigma
;y2 vastaava (tarkista!)
    histo_f,y2,-5.,5.,.05,xx,yy
    nwin
    plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',xs=1,$
        title='Box-Muller menetelmällä luotuja satunnaislukuja'

;teoreettinen tiheysjakauma
    xteo=-5.+10*findgen(1001)/1000.
    yteo=1./sqrt(2.*pi)*exp(-0.5*xteo^2)
    oplot,xteo,yteo,lines=0,col=2,thick=3

    psdirect,program,ps,/stop
end

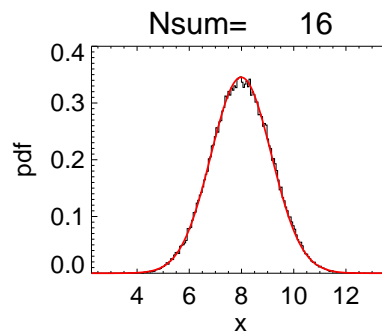
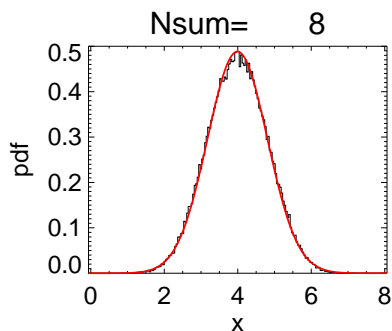
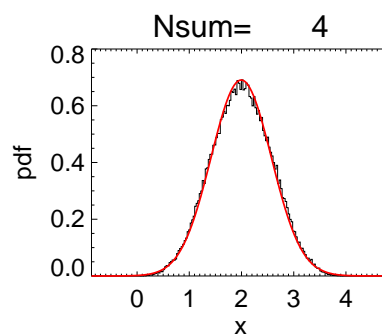
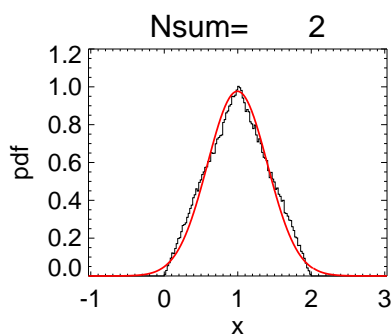
```

4. Normaalijakaantuneita lukuja voidaan tuottaa myös keskeiseen raja-arvolauseeseen pohjautuen. Otetaan $N \gg 1$ tasaisesti jakaantunutta lukua $X_i \sim \text{TAS}(0, 1)$ ja muodostetaan niiden summa. Keskeisen raja-arvolauseen perusteella summan jakauma lähenee normaalijakaumaa $N(\mu, \sigma)$, jossa $\mu =$ summan odotusarvo (nyt $\frac{N}{2}$) ja σ^2 sen varianssin odotusarvo ($\frac{N}{12}$)

Erityisesti,

$$y = \left(\sum_{i=1}^{12} X_i \right) - 6 \sim N(0, 1) \text{ (Miksi?)}$$

Totea käytännössä millaisia normaalijakauman approksimaatioita saat vaihtelemalla summattavien lukujen määriä.



/home/heikki/STATI2008/EXERCISES/ldharj2_kral_tas

heikki@canopus.fysik.yo.oulu.fi Thu Mar 27 09:12:56 2008

Huom: eo arvolla $N = 12$ saatujen lukujen arvoalue on $-6 \leq X \leq 6$. Arvioi kuinka suuri puute tämä on (eli miten suurella todennäköisyydellä oikea normaalijakauma $N(0,1)$ antaisi lukuja taman valin ulkopuolella) Karkean arvion voit tehdä luomalla suuren määrän satunnaislukuja `randomn`-funktion avulla ja katsomalla moniko on tämän välin ulkopuolella.

```
IDL> n=10000001
IDL> x=randomn(seed,n)
IDL> ind=where(abs(x) ge 1, count) & print,count/float(n)
    0.317306
IDL> ind=where(abs(x) ge 2, count) & print,count/float(n)
    0.0456940
IDL> ind=where(abs(x) ge 3, count) & print,count/float(n)
    0.00268100
IDL> ind=where(abs(x) ge 4, count) & print,count/float(n)
    6.10000e-05
IDL> ind=where(abs(x) ge 5, count) & print,count/float(n)
    0.00000
IDL> ind=where(abs(x) ge 6, count) & print,count/float(n)
    0.00000

q;-----
;idlharj2_kral_tas
;-----
;summataan Nsum tasaisesti jaantunutta lukuja
;keskeinen rajarvolause -> summa normaalisesti jakaantunut kun Nsum suuri
;-----
program='idlharj2_kral_tas' & ps=0
psdirect,program,ps,/color

Nsumtab=[2,4,8,16]
!p.multi=[0,2,2]
Notos=1000001

for i=0,n_elements(nsumtab)-1 do begin
    nsum=nsumtab(i)
    xsum=fltarr(notos)
    for j=0,nsum-1 do begin
        xsum=xsum+randomu(seed,notos)
    endfor

;teoreettinen tiheysjakauma
;kaytetaan sopivien plottirajojen arviointiin
    mu=nsum/2.
    sigma=sqrt(nsum/12.)
    xala=mu-5*sigma & xyla=mu+5*sigma & dx=sigma*.05

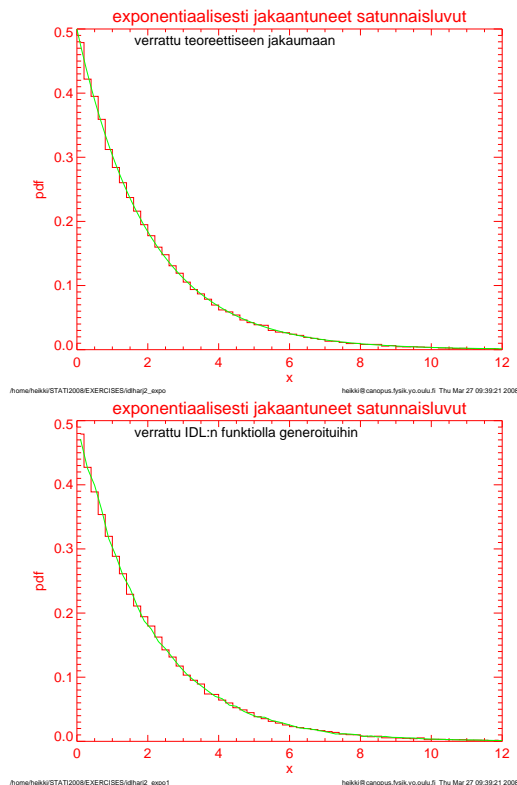
    histo_f,xsum,xala,xyla,dx,xx,yy
    nwin
    plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title='Nsum='+string(Nsum),xs=1
    xteo=(findgen(1001)/1000.*10-5.)*sigma+mu
    yteo=1./sqrt(2.*!pi*sigma^2)*exp(-0.5*(xteo-mu)^2/sigma^2)
    oplot,xteo,yteo,lines=0,col=2,thick=3
endfor
psdirect,program,ps,/stop
!p.multi=0
end
```

5. Luodaan eksponentiaalisesti jakaantuneita satunnaislukuja tasaisesti jakaantuneiden lukujen $R \sim \text{TAS}(0, 1)$ avulla.

$$f(x) = \frac{1}{c}e^{-x/c} \text{ kun } x > 0 \quad (\bar{x} = c)$$

$$\text{Asetetaan } R_i = \int_0^{X_i} f(x)dx \Rightarrow \boxed{X_i = -c \log R_i}$$

Totea että menetelmä toimii, vertaamalla sitä IDL:n `randomu`-funktion avulla luotuihin eksponentiaalisesti jakaantuneisiin lukuihin (kts `?randomu`), sekä teoreettiseen tiheysfunktioon.



```

;-----
;idlharj2_expo
;-----
;luodaan eksponentiaalisesti jakaantuneita satunnaislukuja
;tasaisen jakauman avulla
;-----
program='idlharj2_expo'
ps=0
psdirect,program,ps,/color
N=100000
c=2 ; eksponentiaalijakauman keskiarvo

;luodaan jakauma
r=randomu(seed,N)
x=-c*log(r)

```

```

;histogrammi
  histo_f,x,0.,6*c,0.1*c,xx,yy
  nwin
  plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title='exponentiaalisesti jakaantuneet satunnaisluvut',col=2

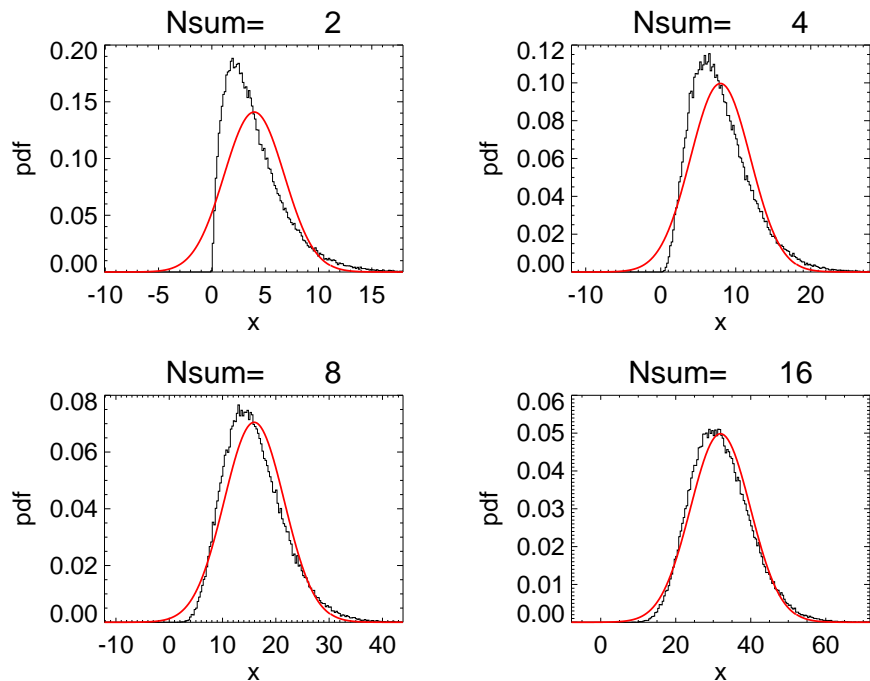
  xyouts,0.25,0.9,/normal,'verrattu teoreettiseen jakaumaan'
  xteo=findgen(1001)/1000.*10*c
  yteo=1./c*exp(-xteo/c)
  oplot,xteo,yteo,lines=0,col=3,thick=3
  psdirect,program,ps,/stop

;-----
;verrataan idl-randomu_function avulla saatuihin
  program='idlharj2_expo1'
  psdirect,program,ps,/color
  N=100000
  c=2
  r=randomu(seed,N)
  x=-c*aolog(r)
  histo_f,x,0.,6*c,0.1*c,xx,yy
  nwin
  plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',col=2,title='exponentiaalisesti jakaantuneet satunnaisluvut'

;idl: funktiolla
  y=randomu(seed,N,gamma=1)*c
  histo_f,y,0.,10*c,0.1*c,xx,yy
  oplot,xx,yy,col=3
  xyouts,0.25,0.9,/normal,'verrattu IDL:n funktiolla generoituihin'
  psdirect,program,ps,/stop
end

```

6. Demonstroi keskeisen raja-arvolauseen pätevyys summaamalla eksponentiaalisesti jakaantuneita satunnaislukuja. Luennoilla annetussa esimerkissä (Wall-Jenkins Kuva 2.6) verrattiin 2, 4, ja 16 eksponenttijakaumasta otetun luvun keskiarvon jakaumia Gaussiseen jakaumafunktion. Käytä näitä samoja arvoja.



/home/heikki/STATI2008/EXERCISES/ldharj2_kral_expo

heikki@canopus.fysik.yo.oulu.fi Thu Mar 27 09:46:50 2008

```

;-----
;ldharj2_kral_tas
;-----
;summataan Nsum eksponentiaalisesti jaantunutta lukuja
;keskeinen rajarvolause -> summa normaalisesti jakaantunut kun Nsum suuri
;-----
program='ldharj2_kral_expo'
ps=0
psdirect,program,ps,/color
!p.multi=[0,2,2]
nwin

Nsumtab=[2,4,8,16]
Notos=1000001
c=2.

for i=0,n_elements(nsumtab)-1 do begin
    nsum=nsumtab(i)
    xsum=fltarr(notos)
    for j=0,nsum-1 do begin
;idl:n funktion avulla
        x1=c*randomu(seed,notos,gamma=1)
        xsum=xsum+x1
    endfor
endfor

```



```

;teoreettinen tiheysjakauma
;kaytetaan sopivien plottirajojen arviointiin
;eksponenttijakauman mu=c, sigma=c
    mu=nsum*c
    sigma=sqrt(nsum)*c
    xala=mu-5*sigma
    xyla=mu+5*sigma
    dx=sigma*.05

    histo_f,xsum,xala,xyla,dx,xx,yy
    plot,xx,yy,psym=10,xtitle='x',ytitle='pdf',title='Nsum='+string(Nsum),xs=1

    xteo=(findgen(1001)/1000.*10-5.)*sigma+mu
    yteo=1./sqrt(2.*!pi*sigma^2)*exp(-0.5*(xteo-mu)^2/sigma^2)
    oplot,xteo,yteo,lines=0,col=2,thick=3
endfor
psdirect,program,ps,/stop
!p.multi=0
end

```

7. Poisson jakauman luonti.

Poisson jakauma kuvaa prosessia, jossa aikayksikköä kohti tapahtuu keskimäärin λ tapahtumaa (raja-arvo binomijakaumasta, kun toistokokeen yksittäisessä yrityksessä on hyvin pieni onnistumistodennäköisyys p , mutta yritysten määrä N on suuri, $\lambda = pN$). Tällöin tapahtuneiden tapausten lukumäärä X noudattaa jakaumaa.

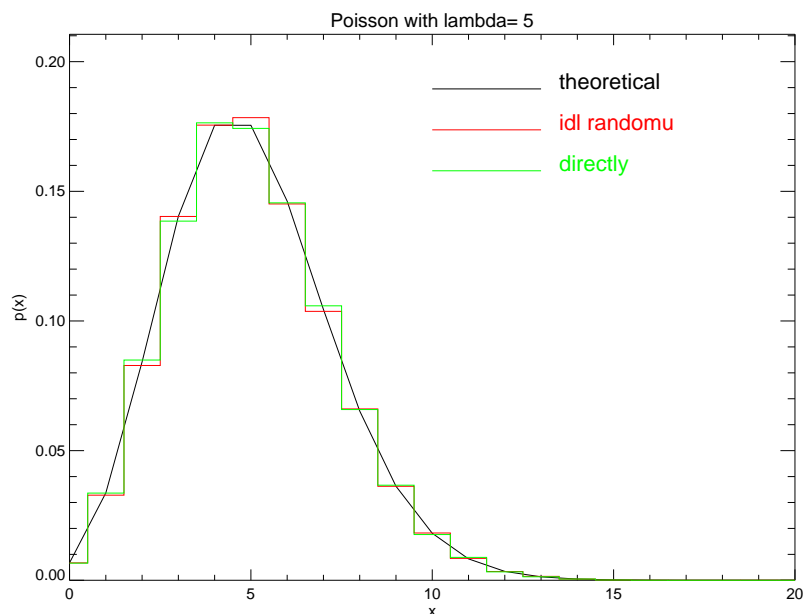
$$p(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, 2, \dots$$

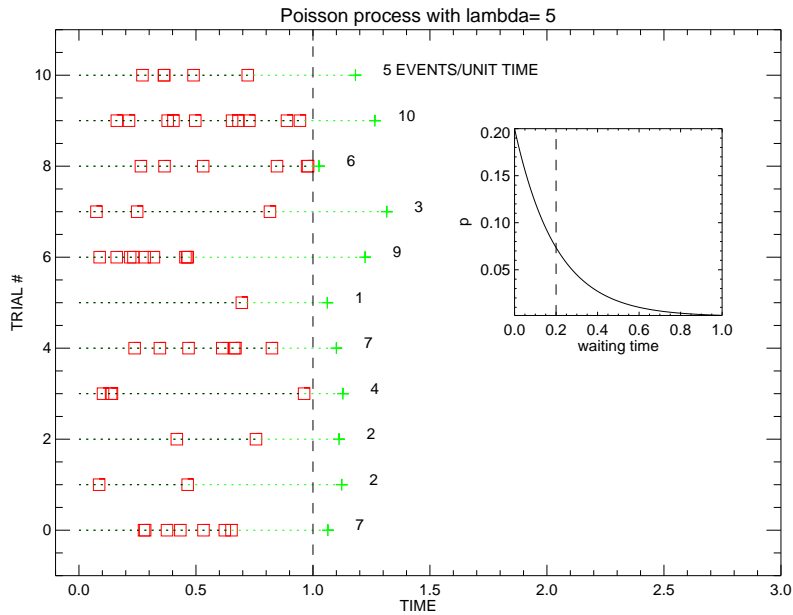
Poisson prosessissa tapahtumien välinen odotusaika on eksponentiaalisesti jakaantunut. Koska tapahtumia on keskimäärin λ /aikayksikkö, niin odotusajan w keskiarvo on $\bar{w} = 1/\lambda$ aikayksikköä. Siten w :n jakauma on

$$f(w) = \frac{1}{\bar{w}} e^{-w/\bar{w}} = \lambda e^{-\lambda w}$$

Poisson jakaantuneita satunnaislukuja voidaan tuottaa summaamalla eo. eksponentiaali-jakauman mukaisesti jakaantuneita odotusaikoja W_i , kunnes niiden summa ylittää aikayksikön. Merkitään summattujen odotusaikojen lukumäärää $X + 1$: // Tällöin tiedetään, että aikayksikön kuluessa tapahtui X tapausta. Näin luotujen X :n arvojen jakauma noudattaa Poisson-jakaumaa.

Tee ohjelma joka tuottaa Poisson jakautuneita satunnaislukuja käyttäen a) IDL randomu-ohjelmaa ja b) edellä kuvattua menetelmää. Vertaa teoreettiseen jakaumaan. Valitse esim. $\lambda = 5$.





```

;-----
;idlhaj2_poisson_rnd
;HS 050308
;-----
program='idlhaj2_poisson_rnd'
ps=0
psdirect,program,ps,/color

;poisson jakauma:
lambda=5
x=findgen(21)
px=exp(-lambda)*lambda^x/factorial(x)

nwin
plot,x,px,xtitle='x',ytitle='p(x)',$
    title='Poisson with lambda='+string(lambda,'i2'),'yr=[0,max(px)*1.2]

;random number using IDL
r=randomu(seed,100000,poisson=lambda)
histo_f,r,-0.5,21,1,xx,yy
oplot,xx,yy,psym=10,col=2

;directly
;Poisson process with lambda events/timeunit
;waiting time between successive events is exponentially distributed,
;with a mean waiting time 1/lambda
;add up X+1 exponentially distributed waiting times, until the sum
;exceed timeunit -> X is the number of events during timeunit
; X is Poisson distributed.

rr=fltarr(100000)
for i=0l,n_elements(rr)-1 do begin
    z=-1./lambda*alog(randomu(seed,20))
    zc=total(z,/cum)          ;cumulative sum
    ind=where(zc gt 1)
    rr(i)= ind(0)             ;number of event during the time interval
                                ;note first  zc > 1  -> ind(0)=0 (x+1=1, x=0)

```

```

;      second zc > 2   -> ind(0)=1 (x+1=2, x=1)
;      etc    zx > x+1 -> ind(0)=x
endfor
histo_f,rr,-0.5,21,1,xx,yy
oplot,xx,yy,psym=10,col=3

coll=1
if(!d.name eq 'PS') then coll=0
label_data,0.5,0.9,['theoretical','idl randomu','directly'],lines=[0,0,0],col=[coll,2,3]
psdirect,program,ps,/stop

;-----
;illustration
;show 11 trials, each event until TIME> UNIT TIME
;number of event poisson distributed
program='idlhaj2_poisson_rnd1'
psdirect,program,ps,/color

nwin
plot,lindgen(2),/nod,xr=[-.1,3],xs=1,yr=[-1,11],ys=1,xtitle='TIME',$
    ytitle='TRIAL #',$
    title='Poisson process with lambda='+string(lambda,'(i2)')

for i=01,10 do begin
    z=-1./lambda*alog(randomu(seed,100))
    zc=total(z,/cum)           ;cumulative sum
    ind=where(zc lt 1,count)
    if(count gt 0) then ind=lindgen(count)
    if(count gt 0) then oplot,zc(ind),zc(ind)*0+i,psym=6,col=2,syms=1
    oplot,zc(count)*[1,1],[i,i],psym=1,col=3

    add=''
    if(i eq 10) then add=' EVENTS/UNIT TIME'

    xyouts,zc(count)+.1,i,string(count,'(i2)')+add
    oplot,[0,zc(count)],[i,i],lines=1,col=3
    if(count gt 0) then oplot,[0,zc(count-1)],[i,i],lines=1
endfor
oplot,[1,1],[-1,11],lines=2

;insert of waiting time distribution
x=findgen(500)/100.*1./lambda
fx=lambda*exp(-x*lambda)
plot,x,fx,pos=[0.65,0.5,0.9,0.8],/noerase,xtitle='waiting time',ytitle='p'
oplot,[1.,1.]/lambda,[0,10],lines=2,col=2

psdirect,program,ps,/color,/stop
end

```