

Suurten lukujen laki

Tutkitaan suurten lukujen lakia käytännössä. Simuloidaan tietokoneella riippumattomia samoinjakautuneita satunnaismuuttujia, jotka saavat todennäköisyydellä 0,5 arvon 0 tai 1. Lasketaan havaintojen keskiarvo ja katsotaan, konvergoiko se kohti odotusarvoa (0,5). Survolla koe voidaan tehdä oheisella koodilla:

```
* PATH=D:\AIKA08\MA08SIMU
* FILE CREATE MA08SIMU
* FIELDS:
* 1 N 8 OBS
* 2 N 8 X
* 3 N 8 CX
* 4 N 8 MX
* END
* FILE INIT MA08SIMU,1000
* VAR OBS=ORDER TO MA08SIMU / OBS is the number of the observation
* VAR X=IND(RND(0)-0.5) TO MA08SIMU
* (RND generates uniformly distributed random variates on (0,1).)
* (IND yields 0 or 1 according to whether the argument is negative or positive.)
* SER CX=CUM(X,1)/OBS TO MA08SIMU
* VAR MX=CX/OBS TO MA08SIMU
* GPLOT MA08SIMU,OBS,MX / LINE=1 YSCALE=0(0.1)1 GRID=1000,0.5 HEADER=_ XLABEL=Havainto
```

Oheisessa kuvassa näkyy selvästi, kuinka keskiarvo konvergoi kohti odotusarvoansa (0,5) havaintojen lukumäärän kasvaessa.

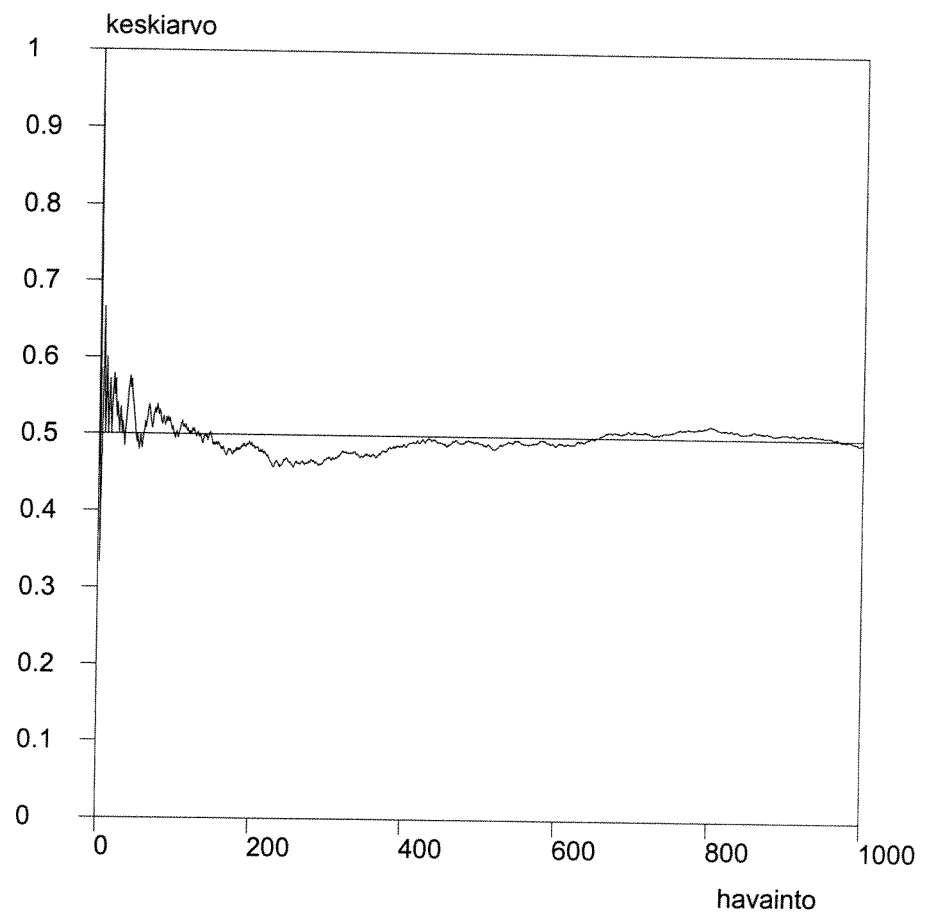


Figure 1:

Keskeinen raja-arvolause

Seuraavat kuusi kuviota ovat David Hendry'n ja Adrian Neale'in kirjasta (1990) *PC-NAIVE. An Interactive program for Monte Carlo experimentation in econometrics* (sivut 76 ja 80). Kirjassa ei valitettavasti kerrota selkeästi, kuinka moneen otokseen kuviot perustuvat. Ilmeisesti kukin kuvio perustuu yli tuhanteen otokseen. Kussakin kuviossa on histogrammi arvotuista satunnaismuuttujista, ja histogrammiin on sovitettu käyrä havainnollistamaan löydettyä pienotosjakaumaa.

Kolmessa ensimmäisessä kuviossa on tietokoneella arvottu tasajakaumasta $Tas(-3, 3)$ riippumattomia havainnoista koostuvia kahden, viiden ja viidenkymmenen kokoisia otoksia. Niistä on kustakin laskettu keskiarvo, niistä piirretty histogrammi ja lopuksi on sovitettu käyrä kuvaamaan estimoitua pienotosjakautumaa. Keskeisen raja-arvolauseen perusteella keskiarvo on normaalijakautunut, kun otos on tarpeeksi suuri. Tasajakauman kohdalla jo kahden kokoinen otos tuottaa melkolailla normaalijakaumaa muistuttavan käyrän. Otokseen kasvaessa pienotosjakauma approksimoi yhä paremmin normaalijakaumaa. Huomaa, että kuvioissa vaaka-akselin skaala pienenee ensimmäisestä kolmanteen, mikä heijastaa keskiarvon varianssin pienenemistä otokseen kasvaessa.

Neljännessä kuviossa havainnollistetaan eksponenttijakaumaa $Exp(2)$. Jakauma on hyvin vino oikealle, joten keskiarvo $Exp(2)$ -jakaumasta arvotuista havainnoista oletettavasti konvergoi hitaammin normaalijakaumaan kuin edellisessä esimerkissä. Viides ja kuudes kuvio havainnollistavat asiaa. Huomaa, että näissäkin kuvissa vaaka-akselin skaala pienenee.

Kolme seuraavaa kuviota ovat Russell Davidsonin ja James MacKinnonin kirjasta (2004) *Econometric theory and methods* (sivu 135 ja 150). Ensimmäisessä kuviossa havainnollistetaan χ^2 -jakauman muotoa eri vapausasteilla. Havaitaan, että $\chi^2(1)$ -jakauma on äärimmäisen vino oikealle.

Seuraavissa kahdessa kuviossa havainnollistetaan jälleen keskeisen raja-arvolauseen toimintaa. Eri kokoisia otoksia on taas arvottu; jälleen ei kerrota toistojen lukumäärää. (Se lienee hyvin suuri, koska havaittuihin jakaumiin sovitetuissa käyrissä ei näy merkkejä satunnaisvaihtelusta.) Kuvioissa esitetään satunnaismuuttujan $\sqrt{n}(\bar{y} - \mu)/\sigma$ jakauma eri havaintomäärillä. Edellä n on havaintojen lukumäärä ja μ ja σ ovat odotusarvo ja varianssi jakaumalle, josta havainnot on arvottu.

Ylemmässä kuviossa havainnollistetaan (jälleen) tasajakaumasta $Tas(0, 1)$ arvottujen havaintojen keskiarvon jakauman konvergointia normaalijakaumaa kohti. Konvergenssi vaikuttaa tavattoman nopealta.

Alemmassa kuviossa keskiarvo on laskettu $\chi^2(1)$ -jakaumasta arvotuista havainnoista. Konvergenssi on huomattavasti hitaampaa. Kun otoskoko on sata, on normaalijakauma approksimaatio kohtalainen muttei läheskään yhtä hyvä kuin edellisessä esimerkissä jo neljän havainnon otoksen kohdalla. Se on intuitiivista, koska tässä esimerkissä havainnot on arvottu äärimmäisen vinosta jakaumasta mutta edellisessä esimerkissä symmetrisestä jakaumasta.

Figure 5 Histogram of sample means of size 2 for $U \sim U[-3,3]$

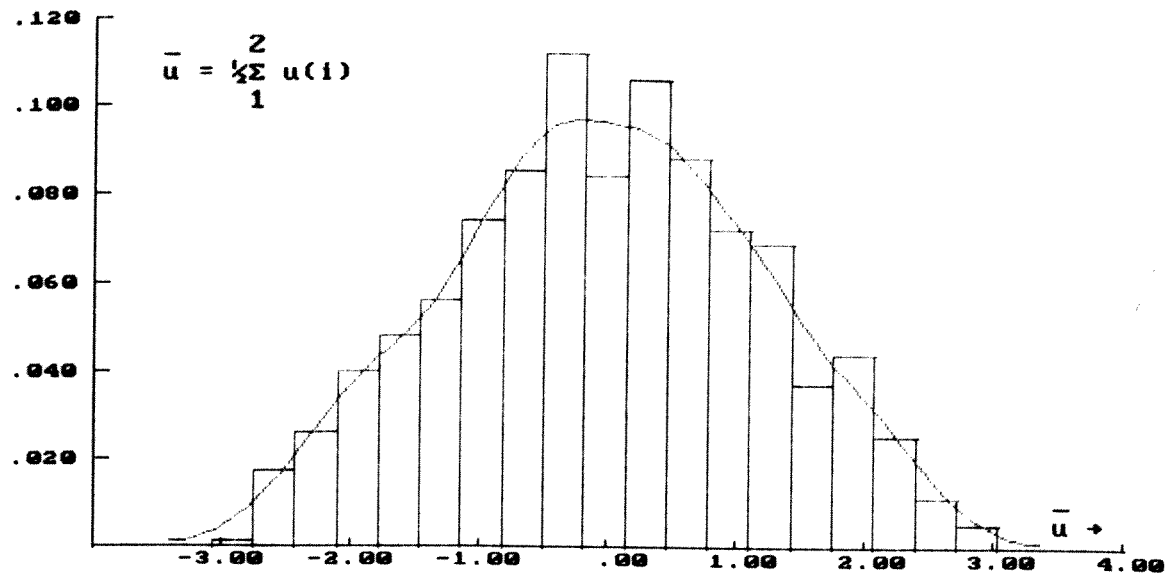


Figure 6. Histogram of sample means of size 5 for $U \sim U[-3,3]$

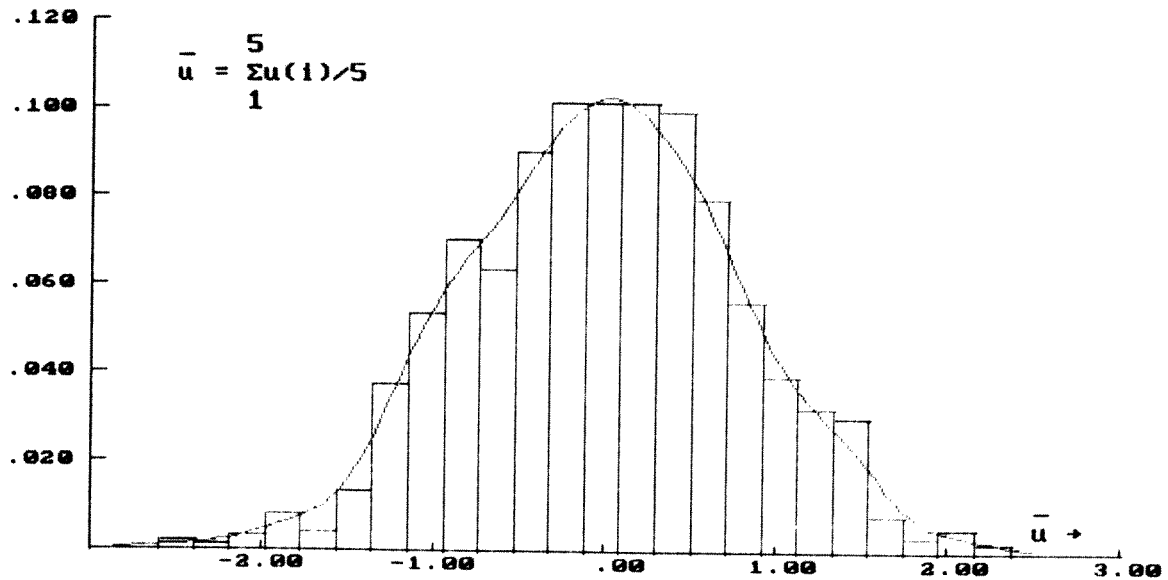


Figure 7. Histogram of sample means of size 50 for $U \sim U[-3,3]$

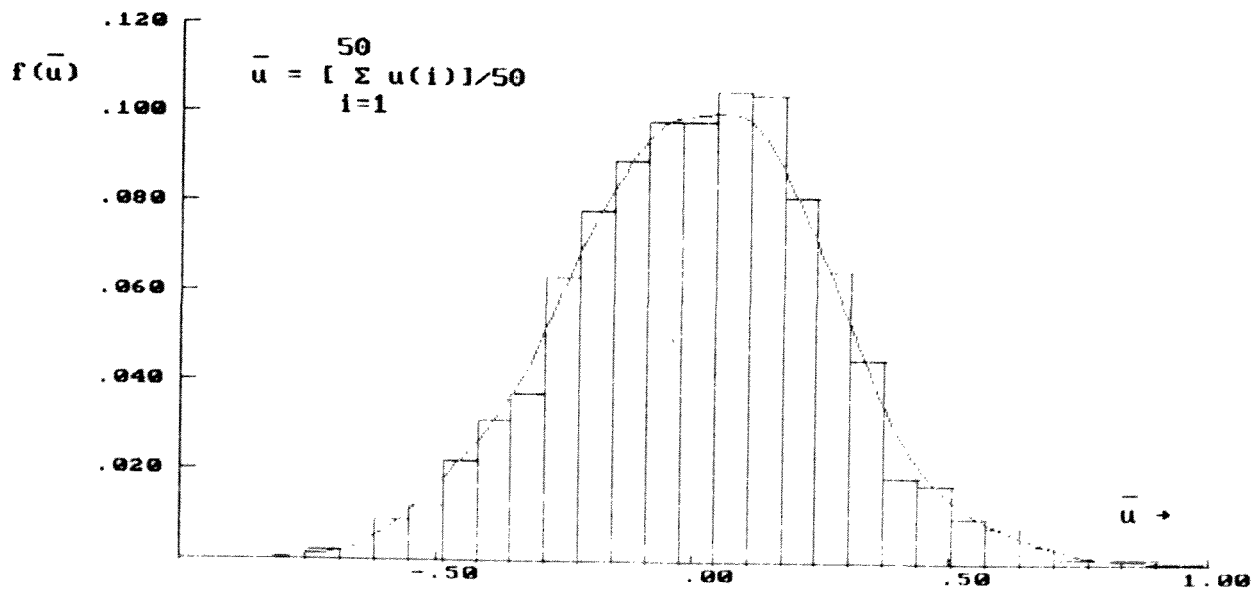


Figure 12. Distribution of $X \sim \text{Exp}(2)$ for single observations.

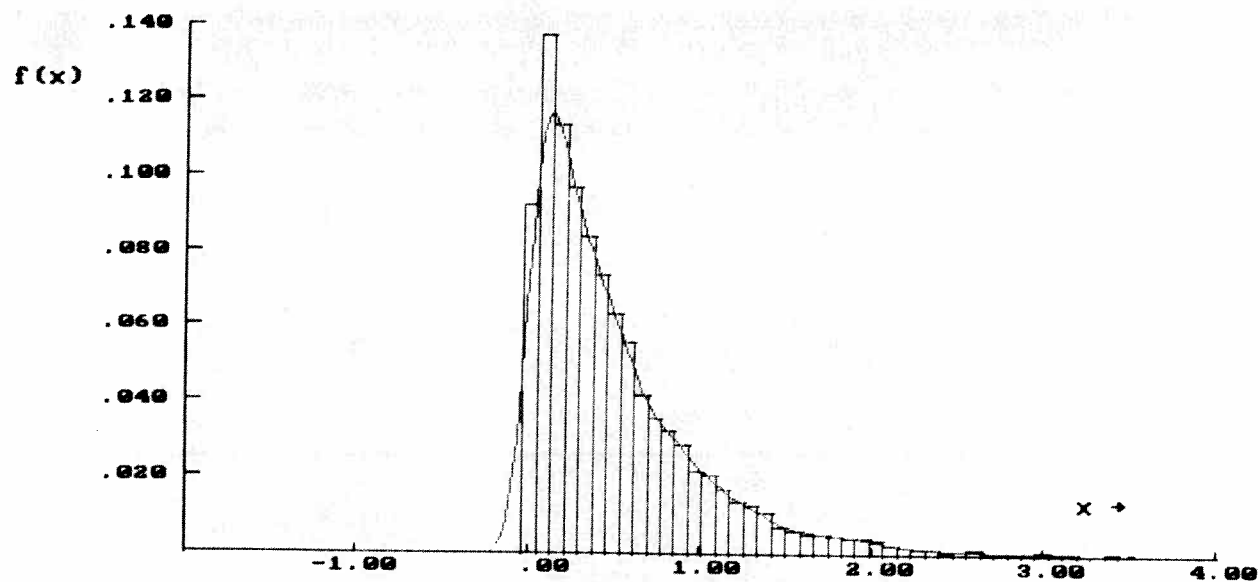


Figure 13. Distribution of sample means of size 5 for $X \sim \text{Exp}(2)$

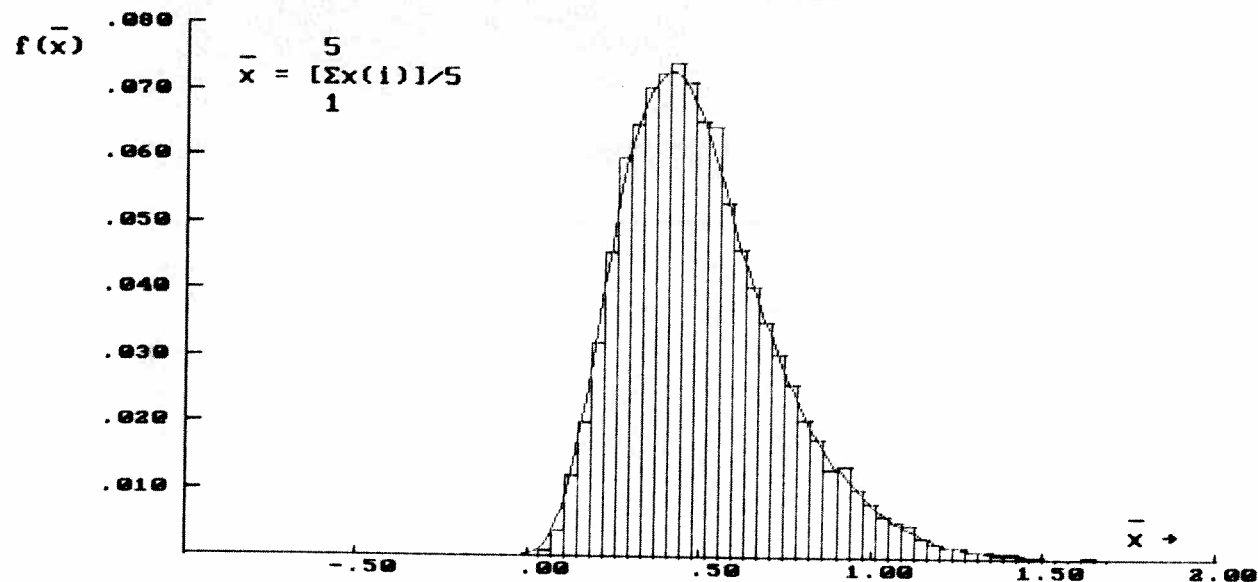
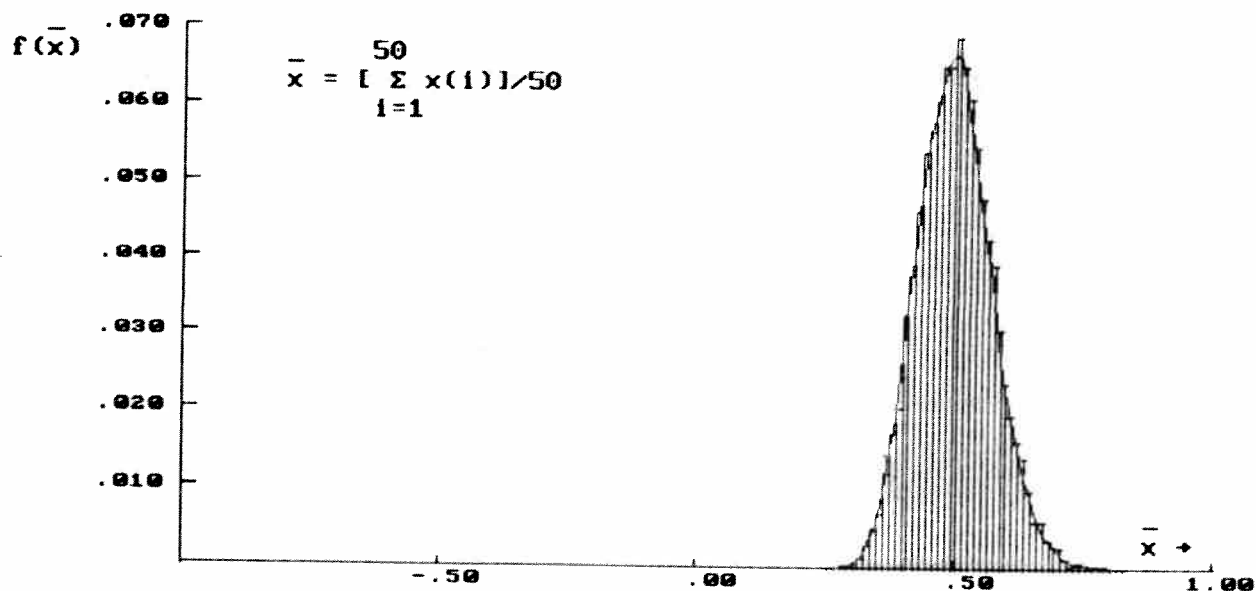


Figure 14. Distribution of sample means of size 50 for $X \sim \text{Exp}(2)$



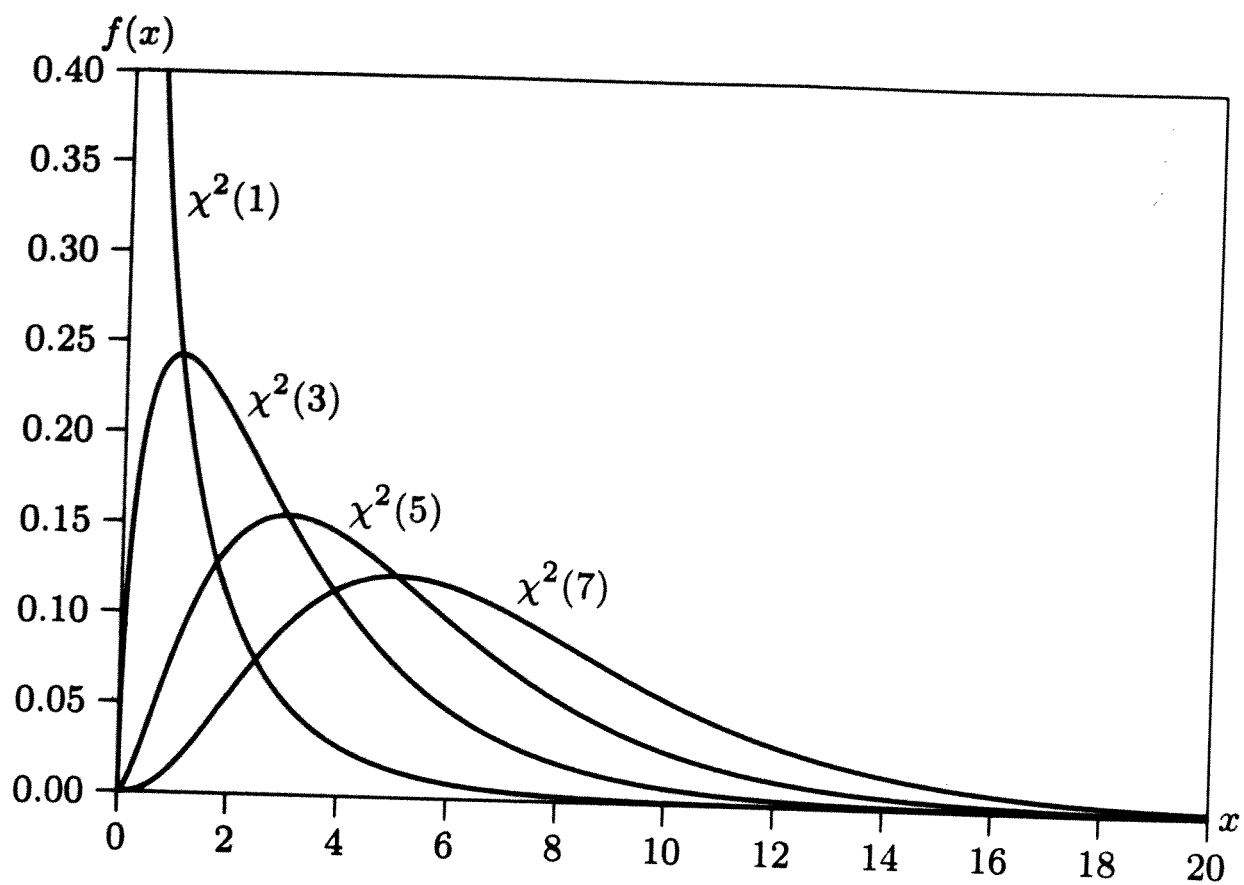


Figure 4.4 Various chi-squared PDFs

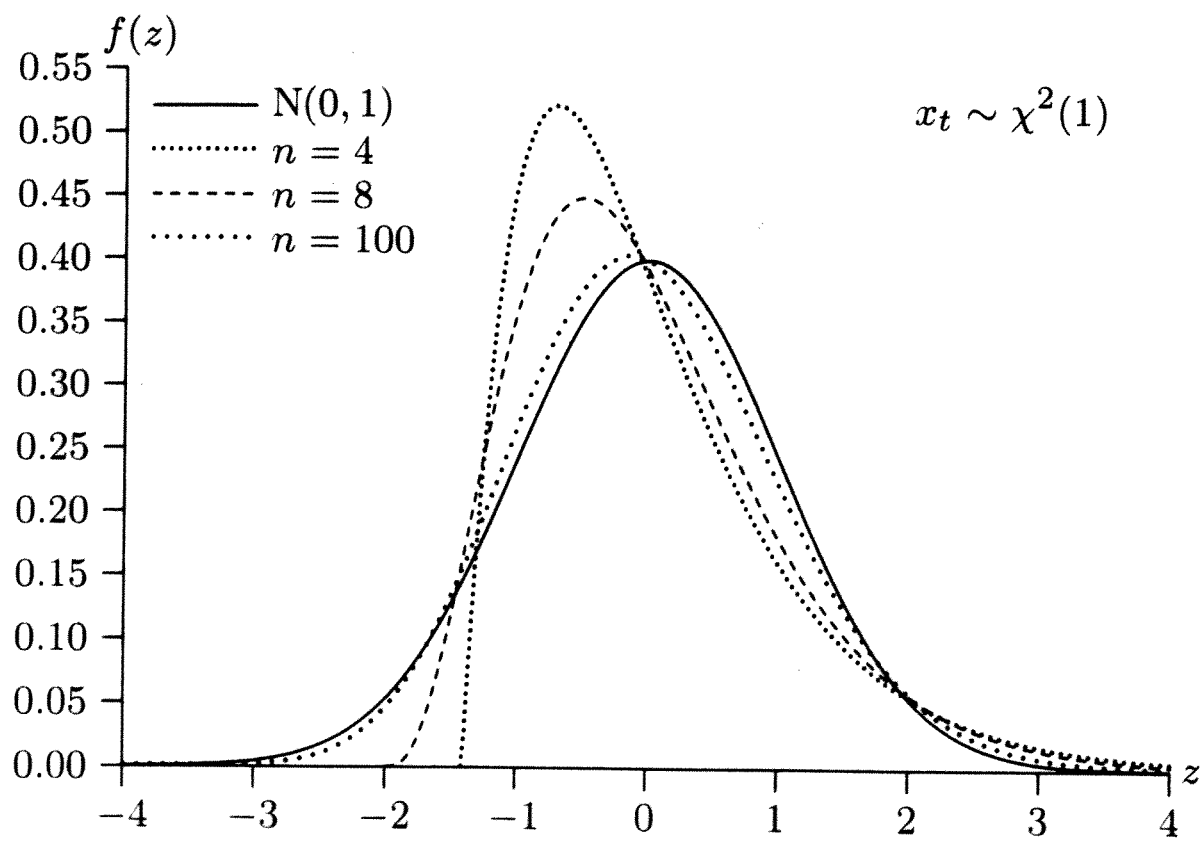
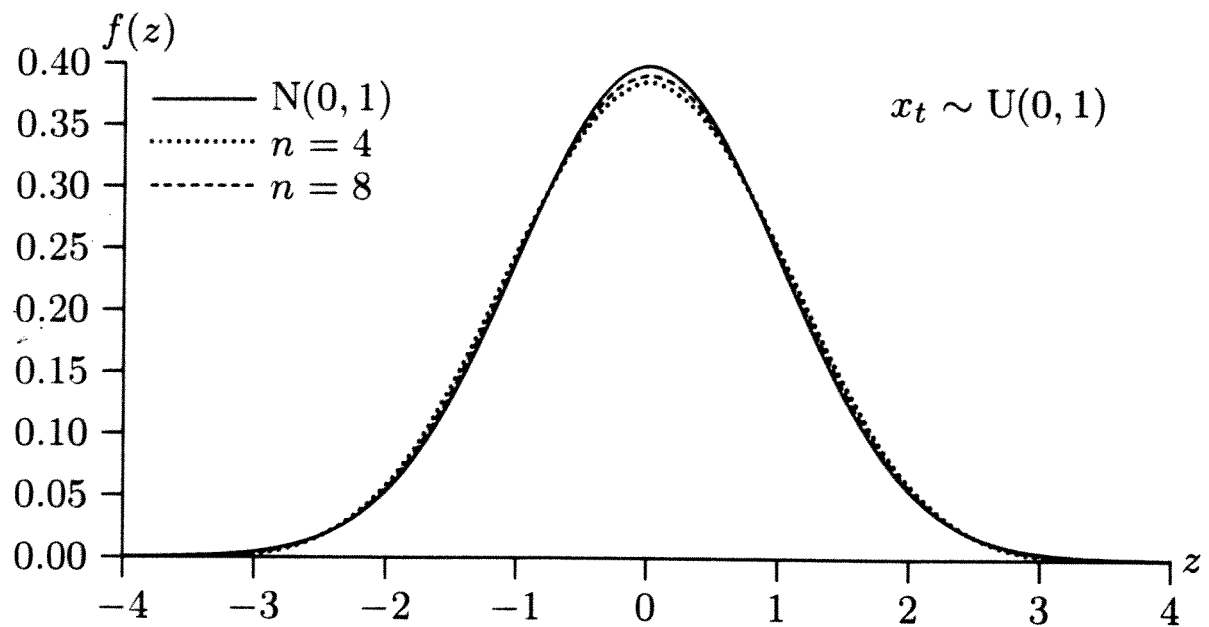


Figure 4.7 The normal approximation for different values of n