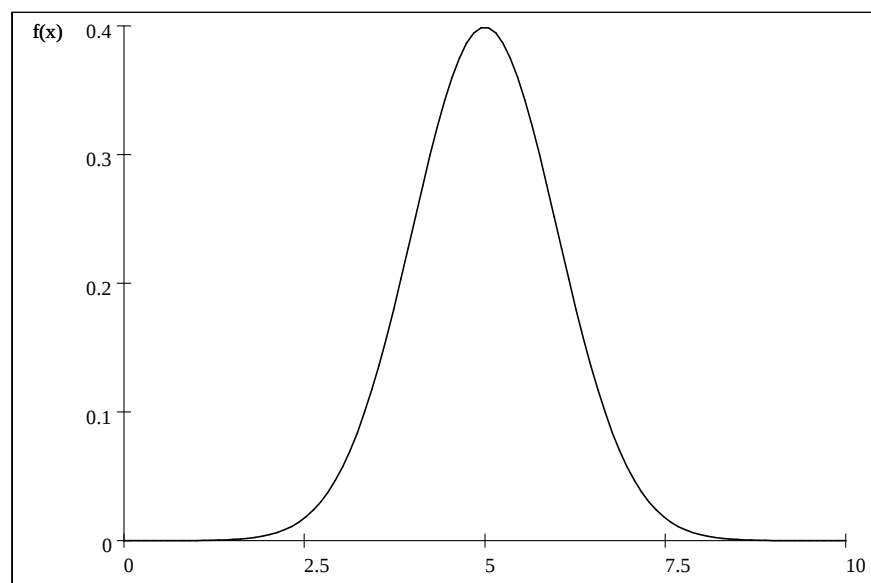

8. NÅGRA SPECIELLA KONTINUERLIGA SANNOLIKHETSFÖRDELNINGAR

8.1 Normalfördelningen

Den kanske viktigaste och mest kända sannolikhetsfördelning är den så kallade normalfördelningen. Den har en mycket stor betydelse som modell för många fenomen och används flitigt inom statistiken. Normalfördelningens täthetsfunktion beskriver en jämn kurva med en bestämd, karakteristisk form, ”klockformen”, se figur 8.1

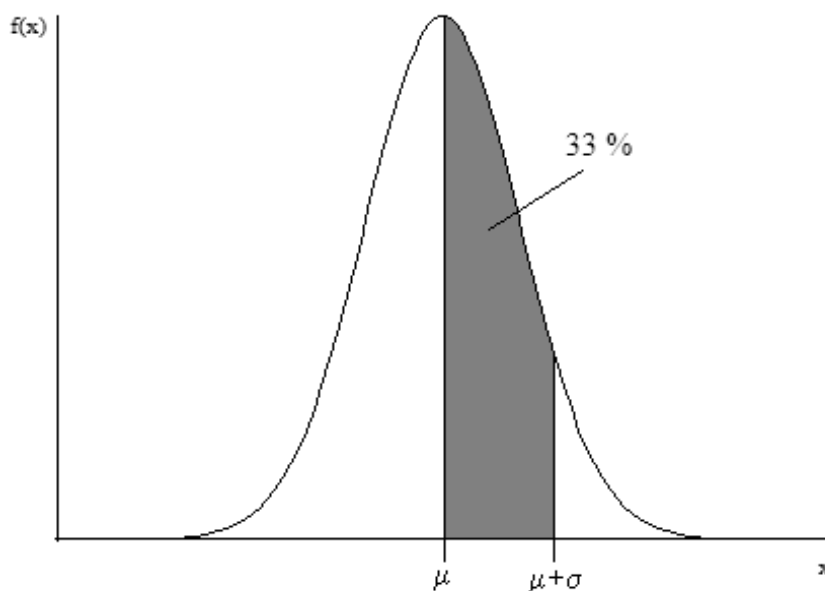


Figur 8.1 Normalfördelningens täthetsfunktion

Allmänt gäller för normalfördelningens täthetsfunktion att den är symmetrisk kring ett mittvärde. Detta mittvärde är fördelningens väntevärde. Normalfördelningens väntevärde brukar symboliseras med den grekiska bokstaven

μ (my). Fördelningens varians brukar symboliseras med σ^2 (sigma kvadrat). Standardavvikelsen är alltså σ . De så kallade svansarna går asymptotiskt mot x -axeln, vilket innebär att kurvan närmar sig x -axeln utan att skära den.

Arean mellan kurvan för täthetsfunktionen och x -axeln representerar totala sannolikheten, som givetvis är lika med 1. Mellan μ och $\mu + \sigma$, där σ är fördelningens standardavvikelse, ligger ungefär 34 % av fördelningsarean, se figur 8.2



Figur 8.2 För normalfördelningen är sannolikheten ungefär 0,34 att få en observation mellan μ och $\mu + \sigma$

Matematiskt uttryckt är normalfördelningens täthetsfunktion

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Den matematiskt insatte inser att kurvan är entydigt bestämd om fördelningens väntevärde och varians är kända storheter. Man använder ofta beteckningen

$X \sim N(\mu, \sigma^2)$ för att ange att en stokastisk variabel X är normalfördelad med väntevärdet μ och variansen σ^2 . Om μ och σ^2 är kända, exempelvis $\mu = 1$ och $\sigma^2 = 4$ skriver vi $X \sim N(\mu = 1, \sigma^2 = 4)$ för att beteckna att X är en normalfördelad stokastisk variabel med väntevärdet 1 och variansen 4. Om det inte finns någon risk för missförstånd skriver vi kortare $X \sim N(1, 4)$.

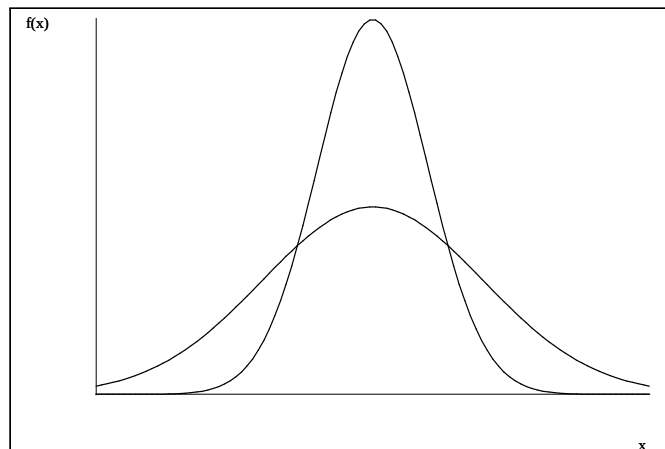
Två normalfördelningar som har samma väntevärde och samma varians är således identiska. Figur 8.3 a) illustrerar två normalfördelningar med samma väntevärde men med olika varianser, medan figur 8.3 b) illustrerar två normalfördelningar med olika väntevärden men med samma varianser.

Normalfördelningen är en teoretisk konstruktion. Inom de tillämpade vetenskaperna existerar knappast några fördelningar som överensstämmer med normalfördelningen i strikt matematisk mening. Däremot kan normalfördelningen ofta med fördel användas som en approximation till andra fördelningar. Vi talar då om en approximativt normalfördelad stokastisk variabel. Ett viktigt exempel på detta är att binomialfördelningen under vissa villkor kan approximeras med normalfördelningen. Detta kommer att diskuteras närmare i avsnitt 8.3.

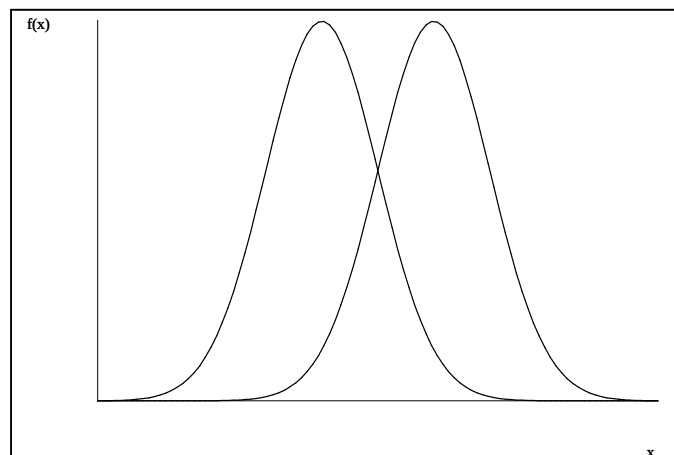
En förutsättning som krävs för vissa inferensmetoder är att de stokastiska variabler som vi arbetar med är normalfördelade eller åtminstone approximativt normalfördelade.

Normalfördelningen har stor betydelse inom inferensteorin också därför att de så kallade samplingfördelningarnas form närmar sig normalfördelningens form, under vissa villkor, då stickprovsstorleken ökar. Samplingfördelningarna behandlas i kapitel 15.

Ofta kan det vara av värde att veta hur stor andel av arean under täthetsfunktionen som ligger till vänster om ett visst variabelvärde, t ex värdet x . Vi erinrar oss att arean till vänster om x representerar sannolikheten att variabeln antar ett värde som är mindre än eller lika med x , dvs $P(X \leq x)$.



Figur 8.3 a) Två normalfördelningar med samma väntevärde men med olika varianser

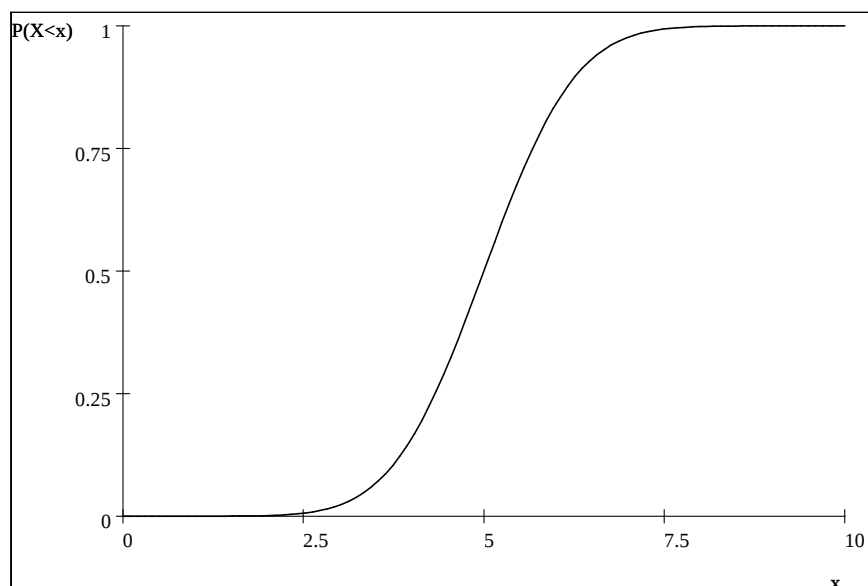


b) Två normalfördelningar med olika väntevärden men med samma varians

För att enkelt kunna avläsa sannolikheten att en normalfördelad stokastisk variabel skall anta ett värde som är mindre än eller lika med ett visst givet tal, kan man använda diagram över normalfördelningens fördelningsfunktion, se figur 8.4. Kurvan beskriver hur andelen av fördelningsarean till vänster om variabelvärdet x ökar då x går från $-\infty$ till $+\infty$. Sålunda avläser vi att för $x = \mu$ är andelen till vänster om x lika med 0,50. Detta innebär att

sannolikheten är 0,50 att en normalfördelad stokastisk variabel skall anta ett värde som är mindre än eller lika med μ , $P(X \leq \mu) = F(\mu) = 0,50$.

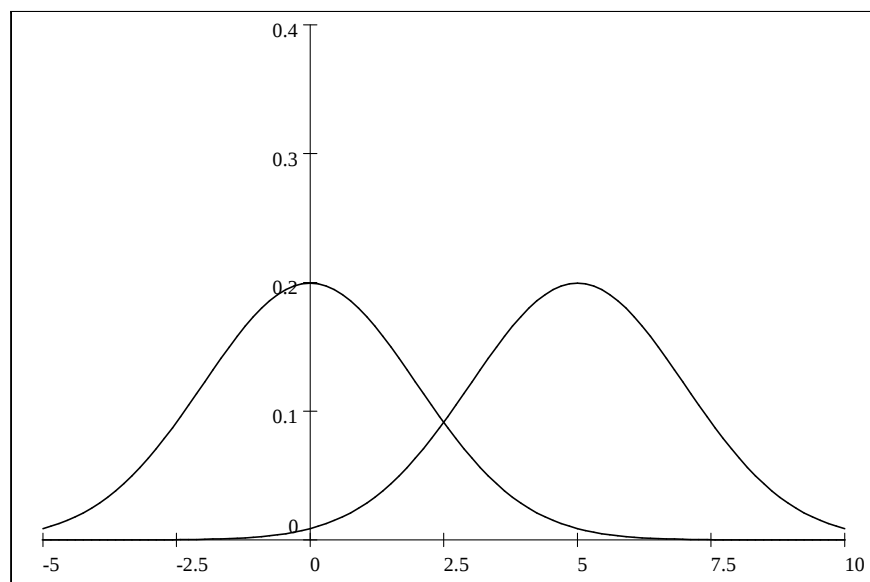
Senare kommer vi att beräkna dessa andelar (sannolikheter) med hjälp av en tabell.



Figur 8.4 Sannolikheten att en normalfördelad stokastisk variabel X antar ett värde som är mindre än eller lika med ett tal x är en funktion av x

8.2 Den standardiserade normalfördelningen

I det här avsnittet studerar vi en speciell variabeltransformation utförd på normalfördelningen. Denna transformaton är ofta av stor nytta då man skall beräkna sannolikheter för normalfördelade stokastiska vriabler. Låt därför X vara en normalfördelad stokastisk variabel med väntevärdet μ och variansen σ^2 . Genom att bilda den stokastiska variabeln $Y = X - \mu$ "flyttar" vi fördelningen så att den blir symmetrisk kring noll i stället för kring väntevärdet μ , se figur 8.5. Med andra ord, fördelningen för Y får väntevärdet 0.



Figur 8.5 Fördelningen blir symmetrisk kring noll då väntevärdet subtraheras från en normalfördelad stokastisk variabel.

Om vi därefter dividerar Y med standardavvikelsen σ , dvs vi bildar en ny stokastisk variabel $Z = \frac{X-\mu}{\sigma}$, finner vi att den nya stokastiska variabeln Z har väntevärdet 0 och standardavvikelsen 1. Denna "Z-transformation" kan vi göra utan att normalfördelningen mister sina allmänna egenskaper.

Då varje normalfördelning entydigt bestäms av sitt väntevärde och sin varians innebär det att varje normalfördelning låter sig transformeras till den entydligt bestämda normalfördelningen med väntevärdet 0 och variansen 1. Denna normalfördelning kallas *den standardiserade normalfördelningen*. En vanlig beteckning för denna är $N(0, 1)$.

Den standardiserade normalfördelningens fördelningsfunktion används så ofta så att den har en egen beteckning. Den betecknas med $\Phi(z)$ och finns tabellerad så att man kan se hur stor andel av fördelningsarean som är belägen till vänster om ett visst z -värde. En förkortad tabell över normalfördelningens fördelningsfunktion visas i tabell 8.1., medan en mer utförlig tabell finns i tabellbilagan.

z	$\Phi(z)$
-3.0	0.001
-2.9	0.002
$\vdots$	$\vdots$
-2,0	0,023
$\vdots$	$\vdots$
-1,0	0.159
$\vdots$	$\vdots$
0.0	0.500
$\vdots$	$\vdots$
0.8	0.788
$\vdots$	$\vdots$
2,0	0,977
$\vdots$	$\vdots$
2.9	0.998
3.0	0.999

Tabell 8.1 Den standardiserade normalfördelningens fördelningsfunktion

Av tabell 8.1 framgår att $\Phi(z) = 0,159$ för $z = -1$, dvs $z = -1$ skär av 15.9 % av fördelningsytan för en standard normalfördelning. Vi kan tolka detta som att sannolikheten att den stokastiska variabeln Z skall anta ett värde mindre än eller lika med -1 är lika med 0.159,

$$P(Z \leq -1) = \Phi(-1) \approx 0,159.$$

Vi skall nu ge några exempel på beräkningar av sannolikheter i samband med normalfördelningen.

Exempel 8.1

Vid det slumpmässiga försöket att bestämma vikten i gram av en viss typ av konserver föreligger en normalfördelning med väntevärdet $\mu = 300$ g och standardavvikelsen $\sigma = 10$ g. Vi söker sannolikheten för att en slumpmässigt

vald konserv har en vikt som understiger 320 g. För att beräkna denna sannolikhet använder vi Z -transformationen. x -värdet 320 motsvarar z -värdet $\frac{320-300}{10} = 2$. Andelen av fördelningsarean till vänster om 2 är 0,977 (se tabell). Detta kan mer formellt skrivas

$$\begin{aligned} P(X \leq 320) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{320 - 300}{10}\right) \\ &= P(Z \leq 2) = \Phi(2) \approx 0,977. \end{aligned}$$

På samma sätt beräknar vi sannolikheten att vikten hos en slumpmässigt vald konserv har en vikt som understiger 280 g genom

$$\begin{aligned} P(X \leq 280) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{280 - 300}{10}\right) \\ &= P(Z \leq -2) = \Phi(-2) \approx 0,023. \end{aligned}$$

Sannolikheten att konservvikten ligger mellan 280 g och 320 g kan vi sålunda få genom

$$\Phi(2) - \Phi(-2) \approx 0,977 - 0,023 = 0,954. \blacksquare$$

Exempel 8.2

Vi skall nu vända på frågeställningen från föregående exempel och beräkna den vikt x sådan att det är en sannolikhet av 0,995 att konservvikten skall vara mindre än x . Vi får

$$\Phi(z) = 0,995$$

vilket innebär

$$\Phi\left(\frac{x - 300}{10}\right) = 0,995.$$

Men

$$\Phi(2,58) \approx 0,995$$

och alltså är

$$\frac{x - 300}{10} \approx 2,58$$

vilket ger

$$x \approx 325,8. \blacksquare$$

På grund av symmetriskäl inser man att

$$\Phi(-z) = 1 - \Phi(z).$$

Vi har nu diskuterat några av normalfördelningens matematiska egenskaper och visat hur man med hjälp av z-transformationen kan beräkna sannolikheter. Normalfördelningen är en matematisk modell som ofta används inom statistiken för att beskriva olika slumpmässiga försök. Det är naturligt att fråga sig i vilka typer av situationer som normalfördelningen kan vara en användbar modell. En typ av modell där normalfördelningen brukar förekomma är den s k mätfelsmodellen. Den bygger på situationen att mätningar av någon fysikalisk storhet skall utföras. Då vi gör mätningar inträffar det ofta att vi inte erhåller exakt det riktiga värdet. Det mätvärde vi erhåller beror inte bara av det sökta värdet, utan det beror även av en rad slumpmässiga faktorer vilka åstadkommer fel i vår mätning. Differensen mellan det sanna sökta värdet och det uppmätta värdet utgör sålunda ett mätfel och detta mätfel är den sammanlagda effekten av ett stort antal slumpmässiga faktorer som påverkar mätningen. Mätfelet kan alltså betraktas som en stokastisk variabel. Den kan vara positiv eller negativ och dess värde kan vara av varierande storlek. Såvida inga systematiska mätfel förekommer blir det förväntade mätfelet lika med noll. De mätfel som uppkommer på detta sätt antas ofta vara normalfördelade. Om X är det erhållna mätvärdet och μ det sanna värdet kan vi uttrycka detta matematiskt som

$$X = \mu + \epsilon$$

där den grekiska bokstaven ϵ (epsilon) representerar mätfelet och vi antar att mätfelet är normalfördelat med väntevärdet 0 och variansen σ^2 ,

$$\epsilon \sim N(0, \sigma^2).$$

En annan typ av modeller där normalfördelningen ofta kommer till användning är då mätvärden varierar i en population. T ex betraktas kroppslängden hos människor ofta som normalfördelat. Om man då väljer en person slumpmässigt och mäter längden kommer den resulterande stokastiska variabeln att vara normalfördelat.

Övning 8.1

Förklara vad väntevärdet μ och standardavvikelsen σ berättar om en normalfördelning.

Övning 8.2

Förklara hur man beräknar ett z -värde som svarar mot ett värde hos en normalfördelad stokastisk variabel. Vad säger z -värdet?

Övning 8.3

En stokastisk variabel X är normalfördelad $N(\mu = 50, \sigma^2 = 100)$. Beräkna med hjälp av tabell

- a) $P(X \leq 65)$. b) $P(X < 25)$. c) $P(X > 35)$. d) $P(X > 70)$.
e) $P(40 < X < 60)$. f) $P(|X - 50| \leq 20)$. g) $P(|X - 50| > 15)$.

Övning 8.4

Vilka värden i en standard normalfördelning definierar

- a) de största 25 % av populationen. b) de största 75 %. c) de minsta 25 %.
d) de mittersta 90 % av populationen.

Övning 8.5

Antag att taxeringsvärdet hos småhus i ett område följer en normalfördelning med väntevärdet 170 000 kr och standardavvikelsen 300 000 kr. Hur stor andel av småhusen i området har ett taxeringsvärde som överstiger 2 000 000?

Övning 8.6

Antag att kolesterolvärdet hos vuxna kvinnor kan beskrivas av en normalfördelning med väntevärdet 180 mg/dl och standardavvikelsen 25 mg/dl.

- a) Hur stor andel vuxna kvinnor har ett kolesterolvärde över 200 mg/dl? b) Hur stor andel vuxna kvinnor har ett kolesterolvärde mellan 150 mg/dl och 170 mg/dl?

Övning 8.7

Vikten i gram av en viss konservertyp är normalfördelad med standardavvikelsen $\sigma = 10$. Väntevärdet μ kan man reglera genom att justera inställningen hos påfyllnadsmaskinen. Vilket är det minsta tillåtna värdet på μ om man vill att högst 5 % av konserverna får ha en vikt understigande 300 gram?

Övning 8.8

Enligt en modell är den årliga utdelningen hos en aktie i företag A normalfördelad med väntevärdet 6,2 % och standardavvikelsen 10,3 %. Enligt samma modell är utdelningen hos en aktie i företag B normalfördelad med väntevärdet 2,6 % med standardavvikelsen 4,3 %. Bestäm sannolikheten att

- a) En aktien i företag A ger en positiv utdelning. b) En aktie i företag B ger en positiv utdelning. c) En aktien i företag A ge en utdelning som är större än 10 %. d) En aktie i företag A ger större utdelning än en aktie i företag B.

8.3 Normalfördelningen som approximation av binomialfördelningen

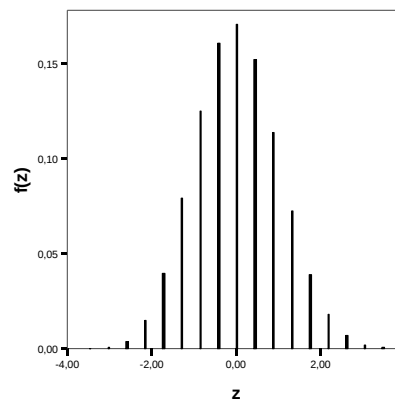
Låt X vara en binomialfördelad stokastisk variabel med parametrarna n och p . Bilda den standardiserade stokastiska variabeln

$$Z = \frac{X - np}{\sqrt{np(1-p)}}.$$

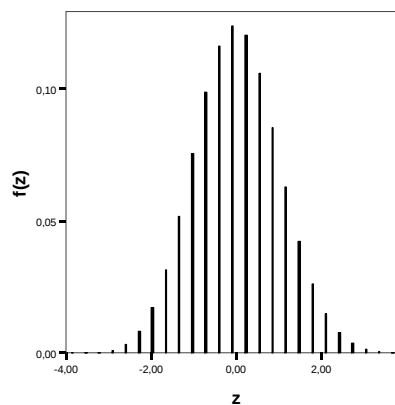
I figurerna 8.6 och 8.7 visas fördelningen för Z då $X \sim \text{bin}(n = 24, p = 1/3)$ respektive $X \sim \text{bin}(n = 48, p = 1/3)$. Från figurerna verkar det som om fördelningen för Z liknar normalfördelningen. Likheten tycks också bli bättre då n ökar. Fördelningen för Z närmar sig den standardiserade normalfördelningen då n ökar. Detta är ett specialfall av den mer generella centrala gränsvärdessatsen som diskuteras längre fram.

Att fördelningen ”liknar” normalfördelningen innebär att normalfördelningen kan användas som approximation till binomialfördelningen under vissa villkor. Approximationen blir bättre då n ökar och är i allmänhet god om $np > 5$

då $p \leq 0.5$ respektive $n(1-p) > 5$ då $p > 0.5$. Tack vare denna approximation är det möjligt att beräkna sannolikheter, åtminstone approximativt, för binomialfördelade stokastiska variabler då n är så stort att fördelningen inte finns tabellerad.



Figur 8.6 Frekvensfunktionen för binomialfördelningen med $n=24$ och $p=1/3$ efter z -transformation



Figur 8.7 Frekvensfunktionen för binomialfördelningen med $n=46$ och $p=1/3$ efter z -transformation

En detalj som komplicerar approximationen är att binomialfördelningen är diskret medan normalfördelningen är kontinuerlig. Vill vi t ex beräkna $P(X = x)$, där x är ett tal, får vi

$$P(X = x) = P\left(\frac{X - np}{\sqrt{np(1-p)}} = \frac{x - np}{\sqrt{np(1-p)}}\right) = P\left(Z = \frac{x - np}{\sqrt{np(1-p)}}\right),$$

men, om Z är en kontinuerlig stokastisk variabel blir denna sannolikhet noll, oavsett värdet på x . Eftersom X är binomialfördelad och endast kan anta heltalsvärden gäller det att $P(X = x) = P(x - \frac{1}{2} \leq X \leq x + \frac{1}{2})$. Vi låter därför $P(X = x)$ approximeras av arean under normalfördelningens täthetsfunktion i ett intervall, se figur 8.8. Med formler får vi

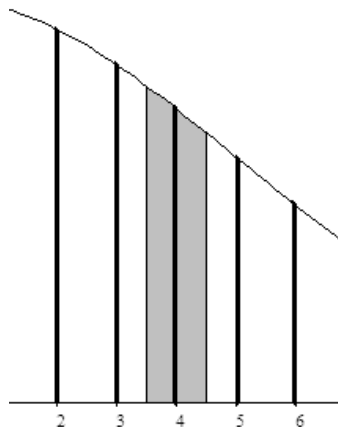
$$\begin{aligned} P(X = x) &= P\left(x - \frac{1}{2} \leq X \leq x + \frac{1}{2}\right) \\ &= P\left(\frac{x - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\ &= P\left(\frac{x - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

Tekniken att på det här sättet bilda ett intervall kring det tal man är intresserad av kallas kontinuitetskorrektion.

Exempel 8.3

Antag att $X \sim \text{bin}(n = 20, p = 0,45)$. Enligt binomialfördelningstabellen gäller det att

$$P(X = 5) = 0,0365.$$



Figur 8.8 $P(X = x)$ i binomialfördelningen approximeras med den skuggade arean i normalfördelningen.

Eftersom $np = 20 \cdot 0,45 = 9 > 5$ kan vi använda normalfördelningen för att approximera denna sannolikhet. Vi får

$$\begin{aligned} P(X = 5) &= \Phi\left(\frac{5 + \frac{1}{2} - 20 \cdot 0,45}{\sqrt{20 \cdot 0,45 \cdot (1 - 0,45)}}\right) - \Phi\left(\frac{5 - \frac{1}{2} - 20 \cdot 0,45}{\sqrt{20 \cdot 0,45 \cdot (1 - 0,45)}}\right) \\ &\approx \Phi(-1,57) - \Phi(-2,02) \approx 0,0582 - 0,0217 = 0,0365, \end{aligned}$$

dvs med fyra decimaler stämmer approximationen exakt.

I det här fallet då den sökta sannolikheten fanns tabellerad är det ingen större idé att arbeta med en approximation om man inte, som här, vill jämföra approximationen med tabellvärdet för att se hur bra approximationen är. Normalt används approximationen då sannolikheterna inte finns tabellerad.

■

Då sannolikheter av typen $P(X \leq x)$ skall beräknas används kontinuitetskorrektionen på motsvarande sätt, dvs

$$P(X \leq x) = P\left(X + \frac{1}{2} \leq x + \frac{1}{2}\right) = P\left(\frac{X + \frac{1}{2} - np}{\sqrt{np(1-p)}} \leq \frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

$$= P\left(Z \leq \frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Övning 8.9

a) Förklara varför det kan vara av intresse att approximera en binomialfördelning med en normalfördelning. b) Vilka villkor måste vara uppfyllda för att en binomialfördelning skall kunna approximeras med en normalfördelning? c) Förklara hur man gör en kontinuitetskorrektur.

Övning 8.10

Antag att $X \sim \text{bin}(n = 80, p = 1/3)$. Beräkna sannolikheterna

a) $P(X = 27)$. b) $P(X \leq 20)$. c) $P(25 \leq X \leq 28)$.

Övning 8.11

En reklamfirma genomförde en kampanj för att introducera en ny produkt. När kampanjen var slutförd hävdade firman att 20 % av målgruppen kände till den nya produkten. För att kontrollera detta drog produktens tillverkare ett slumpmässigt urval om 1000 personer från målgruppen och fann att 150 personer kände till produkten.

a) Antag att reklamfirmans påstående är riktigt. Bestäm förväntat värde och standardavvikelse för antal personer i urvalet som känner till produkten. b) Bestäm sannolikheten att 150 personer eller färre i ett urval om 1000 personer känner till den nya produkten. c) Är det rimligt för tillverkaren att tro på reklamfirmans påstående? Förklara!

8.4 Exponentialfördelningen

Antag att antalet gånger en viss händelse inträffar i ett tidsintervall är Poissonfördelad med parametern λ och låt X vara tiden mellan två sådana händelser. Man kan då visa att X är en kontinuerlig stokastisk variabel med täthetsfunktionen

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{för } x \geq 0 \\ 0 & \text{annars.} \end{cases}$$

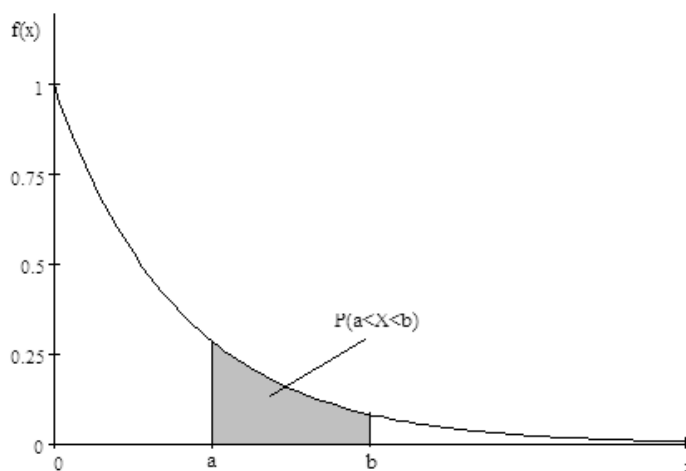
En sådan fördelning kallas exponentialfördelningen med parametern λ . En graf över täthetsfunktionen för en exponentialfördelning visas i figur 8.9. Fördelningsfunktionen för en exponentialfördelad stokastisk variabel är

$$F(X) = 1 - e^{-\lambda x}.$$

Om X är en exponentialfördelad stokastisk variabel med parametern λ betecknar vi det vanligen med $X \sim \exp(\lambda)$. Väntevärdet och variansen i en exponentialfördelning ges av

$$E(X) = \frac{1}{\lambda} \text{ och } V(X) = \frac{1}{\lambda^2}.$$

Tiden mellan två händelser är alltså i medeltal ett delat med antal händelser i ett tidsintervall. Om, t ex, det inträffar i medeltal fyra händelser under en timme är det i medeltal 15 minuter mellan varje händelse.



Figur 8.9. Täthetsfunktionen för en exponentialfördelning.

Exempel 8.4

Antag att antal patienter som kommer till en akutmottagning under en viss tidsperiod är Poissonfördelad. Vi vet då att tiden mellan två ankomster, X , är exponentialfördelad. Det har visat sig att det kommer i medeltal åtta patienter per timme, dvs $\lambda = 8$. Sannolikheten att man får vänta högst tio minuter (dvs $1/6$ timme) mellan två patienter är därför

$$P\left(X \leq \frac{1}{6}\right) = F\left(\frac{1}{6}\right) = 1 - e^{-8/6} \approx 0,7364.$$

Sannolikheten att man får vänta minst fem minuter ($1/12$ timme) är på motsvarande sätt

$$P\left(X > \frac{1}{12}\right) = 1 - F\left(\frac{1}{12}\right) = 1 - (1 - e^{-8/12}) \approx 0,5134.$$

Sannolikheten att det tar mellan fem och tio minuter mellan ankomsten av två patienter är därför

$$P\left(\frac{1}{12} \leq X \leq \frac{1}{6}\right) \approx 0,7364 - 0,5134 = 0,2230.$$

Övning 8.12

Antag att en stokastisk variabel X är exponentialfördelad med $\lambda = 2$.

a) Vilka värden kan X anta? b) Rita grafen av täthetsfunktionen. c) Bestäm $P(X \leq 1)$. d) Bestäm $P(0,25 \leq X \leq 1)$. e) Bestäm $P(X \geq 2)$. f) Beräkna väntevärde och standardavvikelse för X . g) Bestäm sannolikheten att X avviker högst en standardavvikelse från väntevärdet.

Övning 8.13

För ett tillverkningsföretag gäller att antal driftstopp i produktionen är i medeltal två per 500 timmar. Låt X beteckna tiden mellan två driftstopp.

a) Bestäm väntevärdet av X . b) Bestäm sannolikheten att tiden mellan två driftstopp är mellan 100 och 300 timmar.

8.5 χ^2 -fördelningen

Låt $X_1, X_2, \dots, X_n$ vara n stycken oberoende normalfördelade stokastiska variabler, alla med väntevärdet μ och variansen σ^2 . Vi vet då att variablerna

$$Z_i = \frac{X_i - \mu}{\sigma} \text{ för } i = 1, 2, \dots, n$$

är oberoende normalfördelade stokastiska variabler med väntevärdet noll och variansen ett. Bilda en ny stokastisk variabel χ^2 genom att kvadrera de stokastiska variablerna Z_i , dvs beräkna Z_i^2 för $i = 1, 2, \dots, n$ och summera kvadraterna

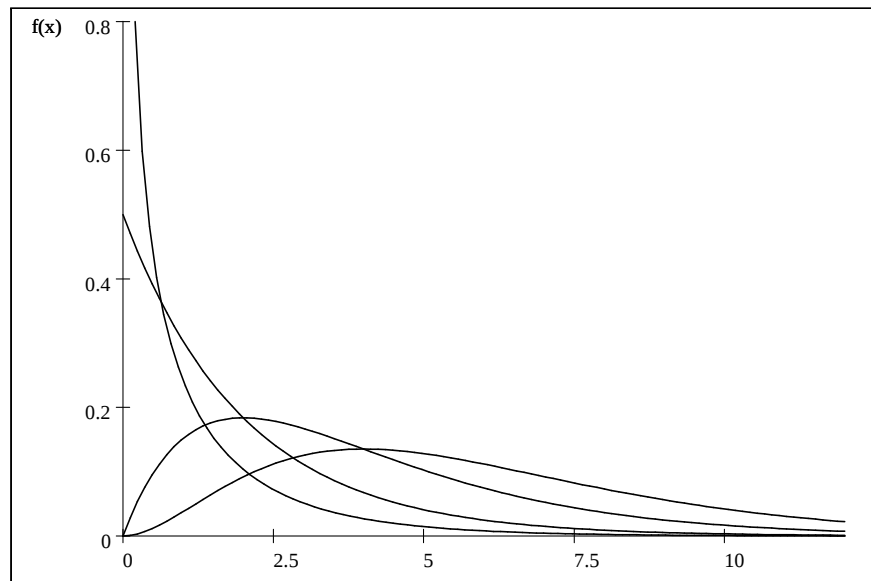
$$\chi^2 = \sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}.$$

χ^2 säges då vara χ^2 -fördelad med n frihetsgrader. Vi betecknar denna fördelning med symbolen $\chi^2(n)$. Antalet termer i summan, n , kallas fördelningens frihetsgradtal.

χ^2 -fördelningen förekommer i flera tillämpningar vi kommer att studera och spelar en viktig roll i inferensteorin. I figur 8.10 visas χ^2 -fördelningens täthetsfunktion vid några olika frihetsgradtal. Observera att den stokastiska variabeln χ^2 inte kan anta negativa värden. Om antalet frihetsgrader n är större än två, så har täthetsfunktionen ett maximum för värdet $n - 2$. Man kan visa att

$$\begin{aligned} E(\chi^2) &= n \\ V(\chi^2) &= 2n. \end{aligned}$$

Då antalet frihetsgrader ökar, blir χ^2 -fördelningen mer symmetrisk. För stora värden på frihetsgradtalet liknar χ^2 -fördelningen normalfördelningen. Vi kan då approximera χ^2 -fördelningen med normalfördelningen. Approximationen blir bättre då frihetsgradtalet ökar. χ^2 -fördelningen finns tabellerad för ett antal frihetsgrader.

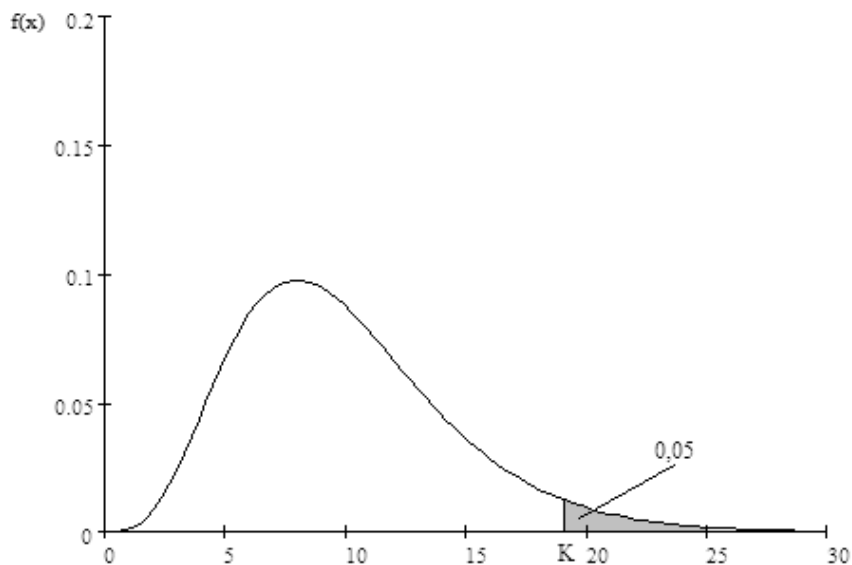


Figur 8.10 χ^2 -fördelningens täthetsfunktion för frihetsgradtalen $n=1$, $n=2$, $n=4$ och $n=6$.

Exempel 8.5

Vi vill bestämma det tal K som är sådant att $P(\chi^2 > K) = 0,05$, där χ^2 är en χ^2 -fördelad stokastisk variabel med 10 frihetsgrader, se figur 8.11.

Eftersom $P(\chi^2 > K) + P(\chi^2 \leq K) = 1$, är vårt problem detsamma som att beräkna det tal K som är sådant att $P(\chi^2 \leq K) = 0,95$. I tabellen ser vi att $P(\chi^2 \leq 18,3) \approx 0,95$. Alltså, vårt sökta tal är $K \approx 18,3$.



Figur 8.11 χ^2 -fördelningen med 10 frihetsgrader. K är det tal som är sådant att $P(\chi^2 > K) = 0,05$

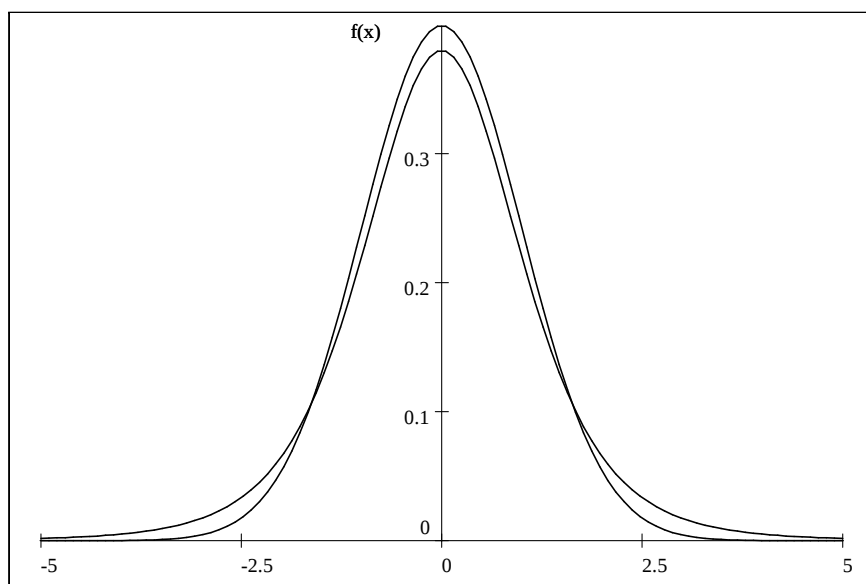
8.6 t-fördelningen

Vi skall betrakta ytterligare en fördelning som är av stor betydelse i tillämpningar av inferensteori. Fördelningen kallas t -fördelningen eller Students t -fördelning. Låt Z vara en standardiserad normalfördelad stokastisk variabel och låt χ^2 vara en χ^2 -fördelad stokastisk variabel med ν frihetsgrader. (ν är den grekiska bokstaven "ny"). Antag att Z och χ^2 är stokastiskt oberoende och bilda en ny stokastisk variabel T genom

$$T = \frac{Z}{\sqrt{\chi^2/\nu}}.$$

Då säges T vara en t -fördelad stokastisk variabel med ν frihetsgrader. Vi betecknar denna fördelning med $t(\nu)$. I figur 8.12 visas täthetsfunktionen för t -fördelningen med $\nu = 5$ frihetsgrader. I figuren visas även täthetsfunktionen för den standardiserade normalfördelningen. De två fördelningarna ser tämligen lika ut. Normalfördelningens täthetsfunktion antar dock större värden nära noll medan t -fördelningens täthetsfunktion har "tjockre svansar". Man kan

visa att då antalet frihetsgrader ökar blir t -fördelningen mer och mer lik den standardiserade normalfördelningen. t -fördelningen finns tabellerad för några frihetsgrader.

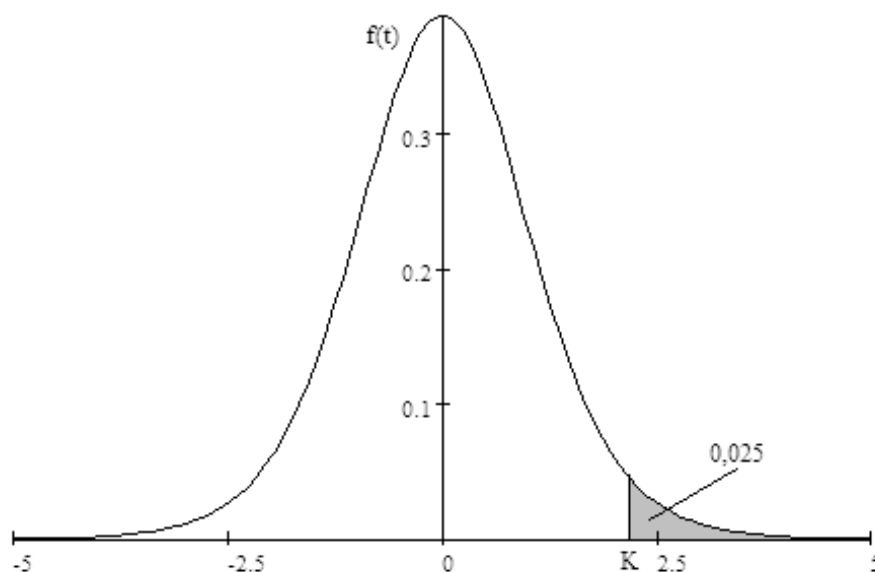


Figur 8.12 Täthetsfunktionerna för t -fördelningen med $\nu = 5$ frihetsgrader och för den standardiserade normalfördelningen

Exempel 8.6

Låt T vara en t -fördelad stokastisk variabel med 10 frihetsgrader (se figur 8.13). Vi vill bestämma ett tal K sådant att $P(T > K) = 0,025$. Vi vet att vårt problem är detsamma som att bestämma det tal K som är sådant att $P(T \leq K) = 0,975$. Ur tabellen får vi att $P(T \leq 2,228) \approx 0,975$. Vårt sökta tal är alltså $K \approx 2,228$.

Om vi i stället studerat en standardiserad normalfördelad stokastisk variabel Z så ger $P(Z > K) = 0,025$ ett annat värde på K , nämligen det som bestäms av $\Phi(K) = 0,975$, dvs $K \approx 1,96$. Det större värdet för t -fördelningen beror tydligen på t -fördelningens "tjockare svansar". ■



Figur 8.13 t -fördelningen med 10 frihetsgrader. K är det tal som är sådant att $P(T > K) = 0,025$

Övning 8.14

En tillverkare av blekmedel önskar reglera sin påfyllnadsmaskin så att i det långa loppet endast en flaska av 100 innehåller mer blekmedel än 2,46 dl. Antag att volymen blekmedel vid en påfyllning kan approximeras med en normalfördelning med standardavvikelsen 0,2 dl. Vad måste väntevärdet vara för att tillverkarens önskemål skall vara uppfyllda?

Övning 8.15

En förpackningsmaskin fyller i medeltal 5,2 hg kaffe i halvkilospaket. Genom erfarenhet vet man att vikten av kaffe i en burk är approximativt normalfördelad med standardavvikelsen 10 g.

- Hur många burkar i ett pari om 100.000 burkar förväntas innehålla mindre än 1/2 kilo kaffe?
- Vad måste medelvikten vara om i medeltal endast en burk av 500 innehåller mindre än 1/2 kg kaffe?