6. STOKASTISKA VARIABLER

6.1 Stokastiska variabler

Om vi tar en lott i ett lotteri med penningvinster är vi givetvis intresserade av om vi vinner eller inte. I det föregående kapitlet har vi diskuterat problem av typen ”hur stor är sannolikheten att vi ska vinna”. Men det finns även andra saker vi är intresserade av vid lottköp. ”Hur mycket vinner vi om vi vinner” och ”hur mycket vinner vi i medeltal om vi köper en lott varje månad” är två frågor vi kan vara intresserade av.

Vid kontroll av belysningen hos bilar är vi kanske mer intresserade av hur många bilar som hade juste belysning än av sannolikheten att en slumpmässigt utvald bil har juste belysning. Vid livslängdsprov av en glödlampa är vi intresserade av hur lång tid lampan fungerar etc.

I dessa exempel är vi intresserade av en variabel, vars värde (lotterivinst, antal bilar med juste belysning, glödlampans livslängd) bestäms av utfallet av ett slumpmässigt försök. En sådan variabel kallas för en *stokastisk variabel.* Ibland säger man också slumpvariabel för att tydliggöra kopplingen mellan variabelns värden och utfall från ett slumpmässigt försök.

I detta kapitel skall vi studera stokastiska variabler samt några mått som karaktäriserar dem. På samma sätt som vi skiljer på diskreta och kontinuerliga variabler skiljer vi här på diskreta och kontinuerliga stokastiska variabler.

Stokastiska variabler brukar betecknas med versaler i slutet av alfabetet, X, Y, Z, U, V osv. När det slumpmässiga försöket som definierar en stokastisk variabel har utförts och utfallet har observerats vet vi sålunda värdet på variabeln. Ett observerat värde på en stokastisk variabel är alltså i sig inte slumpmässigt. Det är därför viktigt att skilja på stokastiska variabler och deras observerade värden. Därför brukar observerade värden betecknas med motsvarande gemener, dvs om X är en stokastisk variabel betecknas dess observerade värde med x, om Y är en stokastisk variabel betecknas dess observerade värde med y osv.

Exempel 6.1

Antag att vi utför det slumpmässiga försöket ”kontroll av belysningen på tio bilar”. Låt A vara händelsen att belysningen fungerar på en bil som kontrolleras. Ett testprotokoll kan t ex visa följande resultat

$$A, \overline{A}, A, A, \overline{A}, A, \overline{A}, A, A, A$$

Vid försöket är vi intresserade av hur många bilar belysningen fungerar för. Vi är alltså inte intresserade av följden testresultat som sådan, utan endast antalet A i sekvensen av kontrollresultat. Därför definierar vi en (diskret) stokastisk variabel som antalet bilar med fungerande belysning. Låt X vara symbol för den stokastiska variabeln. X anger alltså hur många bilar som har juste belysning. De värden som X kan anta är 0, 1, 2,..., 10. För det fall som protokollet redovisar fick vi tydlligen utfallet $x=7$, men vid en annan upprepning av försöket (med tio andra bilar) kan vi få ett annat utfall och ett annat värde på X. ■

Exempel 6.2

Betrakta ett försök att bestämma livslängden i timmar för ett visst slags glödlampor. Antag att man vet att utfallsrummet är intervallet [1200, 1800]. Definiera en (kontinuerlig) stokastisk variabel X som livslängden hos en glödlampa. De värden som X kan anta utgör intervallet [1200, 1800]. ■

Det värde som en stokastisk variabel antar bestäms alltså av ett slumpmässigt försök, dvs mot varje utfall i utfallsrummet svarar ett tal, vilket är den stokastiska variabelns värde. Att den stokastiska variabeln antar ett visst tal betraktas då som en händelse. Sannolikheten att den stokastiska variabeln antar ett visst tal bestäms därför på samma sätt som vi bestämmer sannolikheten för en händelse, dvs vi summerar sannolikheterna för de utfall som leder till händelsen.

Vi kan också tänka oss att definiera ett nytt slumpmässigt försök, vars utfallsrum utgörs av de tal som den stokastiska variabeln kan anta. Utfallsrummet tillsammans med motsvarande sannolikheter bildar då, liksom tidigare, en modell av den stokastiska variabeln. *En modell av en stokastisk variabel kallas sannolikhetsfördelning.*

Exempel 6.3

Betrakta försöket "kasta ett mynt tre gånger och notera vilken sida som kommer upp vid varje kast". Utfallsrummet och motsvarande sannolikheter visas i tabell 6.1. Definiera nu en stokastisk variabel X, som "antal gånger vi får krona". Mot utfallet (KL, KL, KL) svarar talet 0 eftersom vi inte får någon "krona" alls i det utfallet. Eftersom utfallet (KL, KL, KL) är det enda utfall som ger värdet 0 på X får vi

$$P(X = 0) = P(\text{KL, KL, KL}) = 1/8$$

X antar värdet 1 om vi får något av utfallen (KR, KL, KL), (KL, KR, KL) eller (KL, KL, KR). Därför är

$$\begin{aligned} P(X=1) &= P((\text{KR, KL, KL}) \cup (\text{KL, KR, KL}) \cup (\text{KL, KL, KR})) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \end{aligned}$$

På motsvarande sätt får vi de övriga sannolikheterna. Vi ser också att utfallsrummet för X är $\Omega_X = \{0, 1, 2, 3\}$. Den högra delen av tabell 6.1 utgör således sannolikhetsfördelningen för den stokastiska variabeln X. ■

Tabell 6.1 Sannolkhetsfördelningen för den stokastiska variabeln "antal krona vid kast med tre mynt" bestäms från sannolikhetsmodellen för försöket "kasta tre mynt"

Utfall	Sannolikhet	x	$P(X = x)$
(KL, KL, KL)	1/8	0	1/8
(KR, KL, KL)	1/8		
(KL, KR, KL)	1/8	1	3/8
(KL, KL, KR)	1/8		
(KR, KR, KL)	1/8		
(KR, KL, KR)	1/8	2	3/8
(KL, KR, KR)	1/8		
(KR, KR, KR)	1/8	3	1/8

Exempel 6.4

Betrakta försöket att kasta två tärningar. Låt X vara summan av prickarna på de två tärningarna. Utfallsrummet för varje tärningskast är $\{1, 2, 3, 4, 5, 6\}$ och den diskreta stokastiska variabeln X kan anta värdena 2, 3, ..., 12. Motsvarande sannolikheter bestäms genom att summera sannolikheter i det försök som genererar variabeln. T ex får vi

$$P(X = 2) = P(1, 1) = 1/36$$

och

$$P(X = 3) = P((1, 2) \cup (2, 1)) = 1/36 + 1/36 = 2/36. \blacksquare$$

Observara att det är möjligt att definiera många olika stokastiska variabler till varje slumpmässigt försök.

Exempel 6.5

Betrakta exempel 6.4. Där definieras den stokastiska variabeln X som pricksumman vid kast med två tärningar. Det visade sig att utfallsrummet för X är $\Omega_X = \{2, 3, \ldots 12\}$. Låt Y vara produkten av antalet prickar på tärningerna. Tydligen är Y också en stokastisk variabel. Utfallsrummet för Y är $\Omega_Y = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$. Ett tredje exempel på en stokastisk variabel definierad från samma försök får vi om vi bildar skillnaden mellan antalet prickar på den tärning vi kastar först och den vi kastar sedan. Om vi kallar den stokastiska variabeln för Z får vi $\Omega_Z = \{-5, -4, \ldots -1, 0, 1, 2, \ldots, 4, 5\}$. $\blacksquare$

6.2 Frekvens- och täthetsfunktionerna

Låt X vara symbolen för en stokastisk variabel. Vi skall nu studera sannolikhten att X antar ett visst värde, t ex hur stor är sannolikheten att X antar värdet 8 om X är antalet bilar med fungerande belysning då vi testar 10 bilar, dvs vad är $P(X = 8)$? För diskreta stokastiska variabler är den så kallade frekvensfunktionen ett användbart hjälpmedel då man söker sannolikheter av det här slaget. Frekvensfunktionen brukar betecknas med $f(x)$ och definieras genom

$$f(x) = P(X = x).$$

Exempel 6.6

Betrakta försöket att kasta två tärningar i exempel 6.4 och låt X vara summan av antalet prickar som tärningarna får. Det gäller då att

$$\begin{aligned} f(2) &= P(X=2) = 1/36 \\ f(3) &= P(X=3) = 2/36 \\ f(4) &= P(X=4) = 3/36 \\ &\vdots \\ f(11) &= P(X=11) = 2/36 \\ f(12) &= P(X=12) = 1/36 \end{aligned}$$

Vi kan givervis använda $f(x)$ för att beräkna alla sannolikheter om X som vi är intresserade av, t ex får vi sannolikheten att X är udda genom

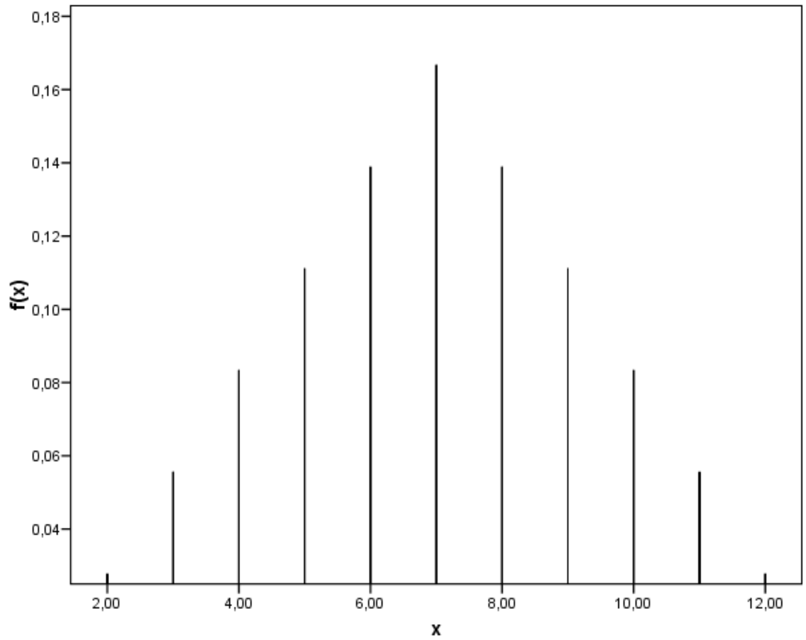
$$\begin{aligned} & f(3)+f(5)+f(7)+f(9)+f(11) \\ =\ & 2/36+4/36+6/36+4/36+2/36 \\ =\ & 18/36 = 1/2. \end{aligned}$$

På motsvarande sätt kan vi beräkna sannolikheten att X ligger i ett visst intervall, exempelvis

$$\begin{aligned} P(4 \leq X \leq 8) &= f(4)+f(5)+f(6)+f(7)+f(8) \\ &= 3/36+4/36+5/36+6/36+5/36 = 23/36. \ \blacksquare \end{aligned}$$

Sannolikhetsfördelningen för en diskret stokastisk variabel kan sammanfattas i en tabel, på det sätt vi har sett i exemplen, eller i ett *stolpdiagram*. Ett stolpdiagram för den stokastiska variabeln ”pricksumman vid kast med två tärningar”, som diskuterades i exemplen 6.4 och 6.6, visas i figur 6.1. I många fall är det även möjligt att ange frekvensfunktionen genom att ange en formel. För pricksumman vid kast med två tärningar kan vi skriva

$$f(x) = \begin{cases} \frac{x-1}{36}, & x = 2, 3, \ldots, 7 \\ \frac{13-x}{36}, & x = 8, 9, \ldots, 12 \end{cases} .$$



Figur 6.1 Frekvensfunktionen för ”pricksumman vid kast med två tärningar”

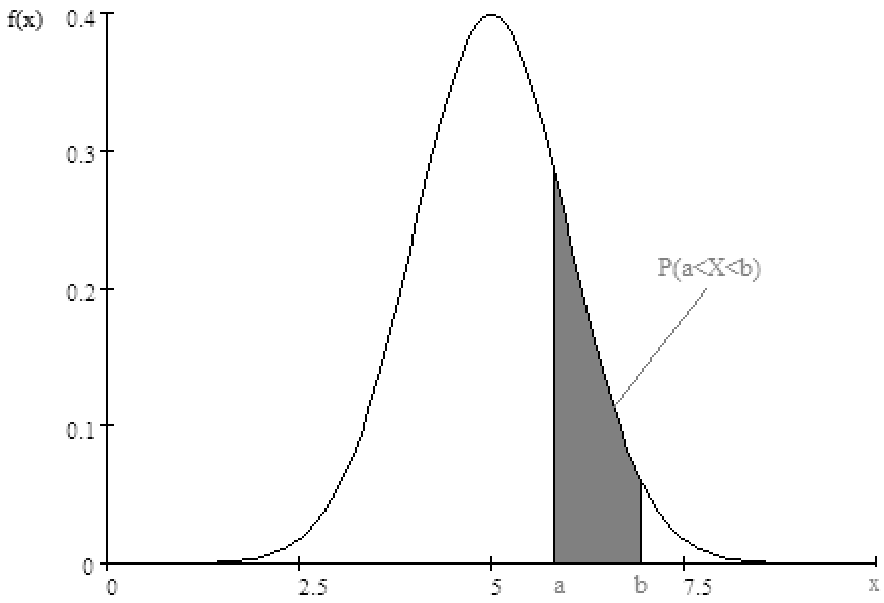
Exempel 6.7

Längden på en slumpmässigt utvald 18-åring är en stokastisk variabel. Beteckna den stokastiska variabeln med X. Om vi mäter längden i hela cm är X diskret och vi kommer att få en viss sannolikhet att X är lika med 180 cm. Beteckna denna sannolikhet med p_1. Mäter vi längden i hela mm, dvs med en större noggrannhet, är X fortfarande diskret. Om p_2 är sannolikheten att X är lika med 1800 mm, så är $p_2 < p_1$ (förklara varför!). Ökar vi precisionen ytterligare kommer motsvarande sannolikhet att bli ännu mindre. För varje ökning i mätnoggrannheten kommer också X att bli mer lika en kontinuerlig stokastisk variabel. Om mätningen görs med ”hur stor precision som helst” får vi att X är en kontinuerlig stokastisk variabel och att sannolikheten att X är precis 180 cm är noll. ■

Av exempl 6.7 framgår att vi inte kan definiera frekvensfunktioner för kontinuerliga stokastiska variabler. Gjorde vi det skulle ju frekvensfunktionen alltid bli noll i det kontinuerliga fallet. I stället för att arbeta med sannolikheter att variabeln antar speciella tal, som i det diskreta fallet, arbetar vi med sannolikheter att variabeln antar värden i ett intervall av tal i det kontinuerliga fallet. Vi definierar därför *täthetsfunktionen* för en kontinuerlig stokastisk variabel så att sannolikheten att variabeln antar ett värde i ett visst intervall är lika med arean under frekvensfunktionen i intervallet, se figur 6.2. Matematiskt kan detta uttryckas som

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

där a och b är intervallgränserna och X är en stokastisk variabel med täthetsfunktionen $f(x)$. ■



Figur 6.2 Sannolikheten att en stokastisk variabel får ett värde i ett intervall (a,b) är lika med arean under täthetsfunktionen.

Täthetsfunktionen kan åskådliggöras med ett diagram som i figur 6.2. Vanligen anges täthetsfunktionen med en formel. En viktig sannolikhetsfördelning är den så kallade normalfördelningen, vilken beskrivs närmare i kapitel 8. Dess täthetsfunktion kan skrivas

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

Med hjälp av frekvensfunktionen respektive täthetsfunktionen kan vi beräkna vilka sannolikheter som helst. Därför kan vi säga att frekvensfunktionen eller täthetsfunktionen fullständigt karaktäriserar sannolikhetesfördelningen. Kunskap om frekvens- eller täthetsfunktionen innebär alltså fullständig kunskap om sannolikhetsfördelningen.

6.3 Fördelningsfunktionen

Vid sidan om frekvens- och täthetsfunktionerna för stokastiska variabler spelar fördelningsfunktionen en stor roll. Vi använder symbolen $F(x)$ för fördelningsfunktionen. Den anger sannolikheten att den stokastiska variabeln skall anta ett värde som är mindre än eller lika med ett tal x, dvs

$$F(x) = P(X \leq x)$$

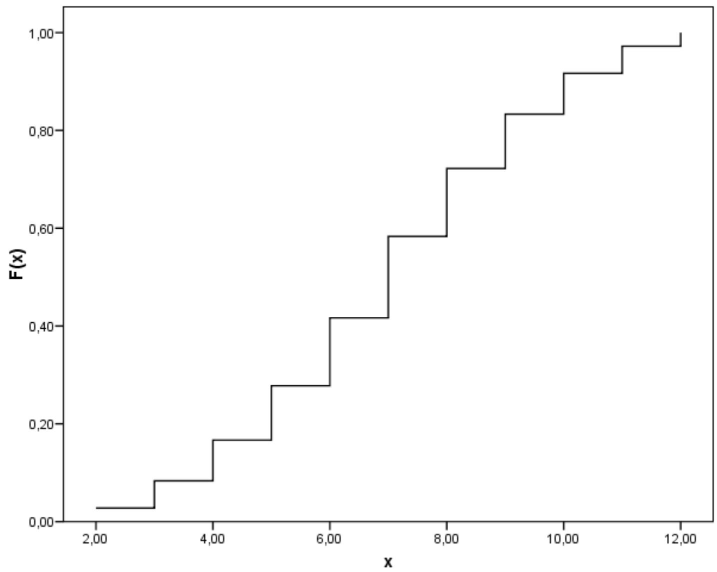
för en stokastisk variabel X. Då X är diskret gäller det därför att

$$F(x) = \sum_{t \leq x} f(t)$$

och då X är kontinuerlig att

$$F(x) = \int_{-\infty}^{x} f(t)\, dt.$$

Frekvens- respektive täthetsfunktionen och fördelningsfunktionen är alltså närbesläktade med varandra. Känner man den ena kan man, såvida man inte stöter på rent matematiska svårigheter, beräkna den andra. För en diskret stokastisk variabel är fördelningen en så kallad trappstegsfunktion, den är konstant utom i de värden som variabeln kan anta värden. Där gör funktionen språng och språngens storlek motsvarar sannolikheten att variabeln skall anta detta värde. Ett exempel visas i figur 6.3 Sambandet mellan täthetsfunktionen för en kontinuerlig stokastisk variabel och fördelningsfunktionen finns illustrerad i figur 6.4.



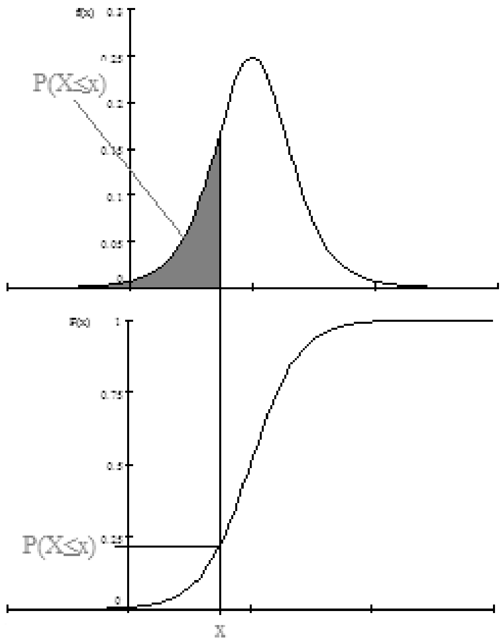
Figur 6.3 Fördelningsfunktionen för en diskret stokastisk variabel är en trappstegsfunktion

Exempel 6.8

Betrakta försöket att kasta två tärningar igen. Tydligen gäller det att

$$\begin{aligned} F(2) &= f(2) = 1/36 \\ F(3) &= f(2) + f(3) = 1/36 + 2/36 = 3/36 \\ F(4) &= f(2) + f(3) + f(4) = 1/36 + 2/36 + 3/36 = 6/36 \\ &\vdots \\ F(12) &= f(2) + f(3) + \ldots + f(12) = 1. \end{aligned}$$

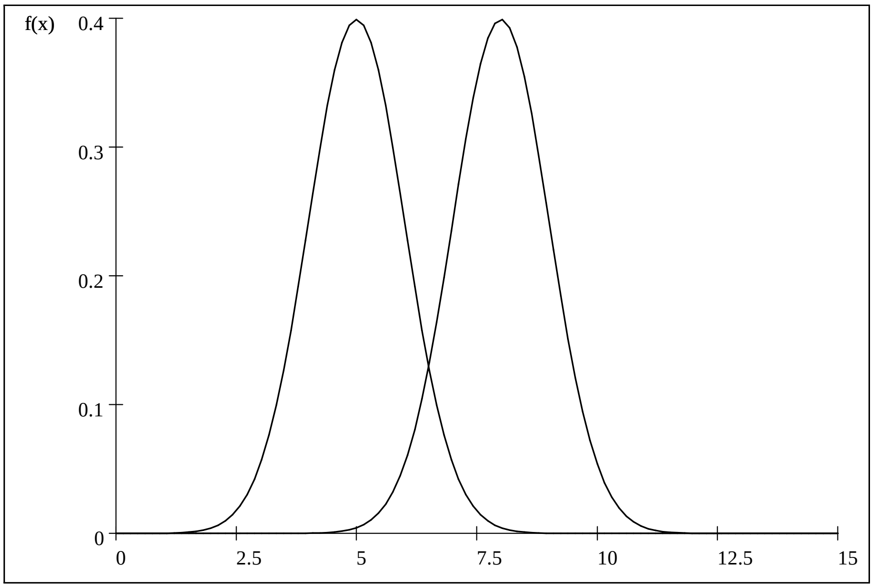
Figur 6.3 visar en figur av denna fördelningsfunktion. ■



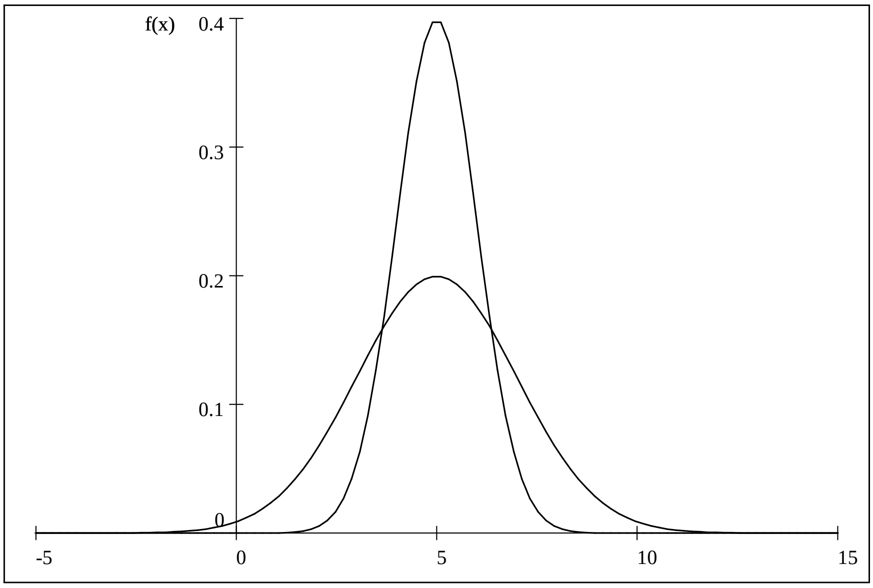
Figur 6.4 Sambandet mellan täthetsfunktionen f(x) och fördelningsfunktionen F(x) för en kontinuerlig stokastisk variabel

6.4 Väntevärde och varians

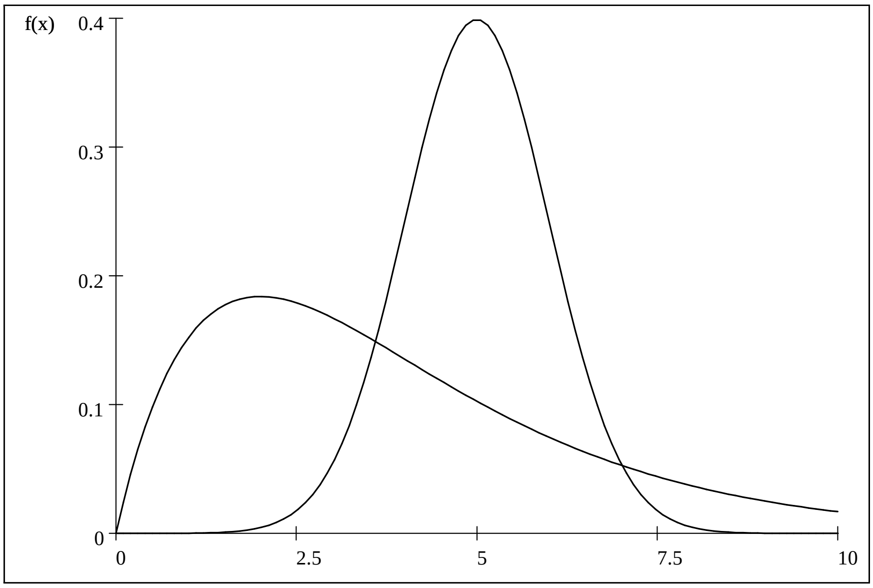
Från avsnitt 6.2 vet vi att frekvens- respektive täthetsfunktionerna för en stokastisk variabel fullständigt karaktäriserar sannolikhetsfördelningen för variabeln. Olika stokastiska variabler kan ha olika sannolikhetsfördelningar och därmed olika frekvens- eller täthetsfunktioner. Olika frekvensfunktioner kan sedan vara skilda åt beträffande läge, som i figur 6.5 eller beträffande spridning, som i figur 6.6. I det här avsnittet skall vi ägna oss åt några enkla mått som just anger läget och spridningen för fördelningar för stokastiska variabler. Givetvis kan olika frekvens- och täthetsfunktioner även ha andra skillnader. De kan t ex vara symmetriska eller sneda, se figur 6.7 för en illustration. Sådana skillnader skall vi inte diskutera här.



Figur 6.5 Två täthetsfunktioner med olika lägen



Figur 6.6 Två täthetsfunktioner med samma läge men med olika spridning



Figur 6.7 En symmetrisk och en sned täthetsfunktionen

Exempel 6.9

Vid ett visst spel kan man vinna 10, 5, 2 eller 0 kronor. Definiera den stokastiska variabeln X som vinsten i ett spel. X kan alltså anta värdena 0, 2, 5 och 10. Vi låter sannolikheterna för de olika utfallen vara 0.4, 0.3, 0.2 och 0.1. Vi kan sammanställa utfallen och sannolikheterna i en tabell på följande sätt

Utfall	x	0	2	5	10
Sannolikhet	$f(x)$	0.4	0.3	0.2	0.1

Antag att spelet spelas ett stort antal gånger, säg 1000 gånger. Vi väntar oss då att de olika utfallen skall inträffa c:a 400, 300, 200 och 100 gånger, dvs vi väntar oss en total vinst av c:a

$$400 \cdot 0 + 300 \cdot 2 + 200 \cdot 5 + 100 \cdot 10 = 2600 \text{ (kronor)}$$

Den förväntade vinsten per spelomgång är alltså

$$\begin{aligned}
& \frac{400 \cdot 0 + 300 \cdot 2 + 200 \cdot 5 + 100 \cdot 10}{1000} \\
= & \frac{400}{1000} \cdot 0 + \frac{300}{1000} \cdot 2 + \frac{200}{1000} \cdot 5 + \frac{100}{1000} \cdot 10 \\
= & 0.4 \cdot 0 + 0.3 \cdot 2 + 0.2 \cdot 5 + 0.1 \cdot 10 = 2.60 \text{ (kronor)}
\end{aligned}$$

Detta är den stokastiska variabelns *väntevärde*. Lägg märke till att vi multiplicerade de olika värdena som den stokastiska variabeln kan anta med motsvarande sannolikheter och adderade de erhållna produkterna. ■

Låt X vara en diskret stolastisk variabel som kan anta värdena $x_1, x_2, \ldots, x_n$. Låt vidare frekvensfunktionen för X vara $f(x)$. Det förväntade värdet eller väntevärdet av X brukar symboliseras med $E(X)$. Formellt definieras väntevärdet genom

$$\begin{aligned}
E(X) &= x_1 f(x_1) + x_2 f(x_2) + \ldots + x_n f(x_n) \\
&= \sum_{i=1}^{n} x_i f(x_i).
\end{aligned}$$

Lägg märke till att väntevärdet för en stokastisk variabel kan tolkas som ett slags medelvärde. Väntevärdet anger därför läget hos frekvensfunktionen.

För en kontinuerlig stokastisk variabel definieras väntevärdet analogt

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

Exempel 6.10

Betrakta det slumpmässiga försöket att kasta två tärningar i föregående avsnitt. Här gäller tydligen

x	$f(x)$	$x \cdot f(x)$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
Summa		252/36

Väntevärdet är alltså 252/36=7, dvs i medeltal förväntar vi oss att få pricksumman 7 då vi kastar två tärningar. ■

Som mått på spridningen används ofta variansen eller standardavvikelsen. Om vi som vanligt låter X vara en stokastisk variabel låter vi $V(X)$ vara symbol för variansen av X. Om X är en diskret stokastisk variabel som kan anta värdena $x_1, x_2, \ldots, x_n$ definierar vi variansen för variabeln genom

$$
\begin{aligned}
& V(X) \\
= \; & (x_1 - E(X))^2 \cdot f(x_1) + (x_2 - E(X))^2 \cdot f(x_2) + \ldots + (x_n - E(X))^2 \cdot f(x_n) \\
= \; & \sum_{i=1}^{n} (x_i - E(X))^2 \cdot f(x_i)
\end{aligned}
$$

Om X är en kontinuerlig stokastisk variabel definieras variansen analogt genom

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\, dx$$

Standardavvikelsen definieras som kvadratroten ur variansen. I figur 6.6 visas täthetsfunktionerna för två kontinuerliga stokastiska variabler, den ena med stor varians, den andra med liten varians.

Vid beräkning av varianser för diskreta stokastiska variabler förenklas ofta arbetet om man utnyttjar följande samband

$$V(X) = \sum_{i=1}^{n} x_i^2 f(x_i) - (E(X))^2 .$$

Exempel 6.11

I exempel 6.9 kan variansen (och även väntevärdet) beräknas enligt följande

x	$f(x)$	$x \cdot f(x)$	$x^2 \cdot f(x)$
0	0.4	0	0
2	0.3	0.6	1.2
5	0.2	1.0	5.0
10	0.1	1.0	10.0
Summa	1.0	2.6	16.2

$$\begin{aligned} E(X) &= 2.6 \\ V(X) &= 16.2 - 2.6^2 = 9.44 \ \blacksquare \end{aligned}$$

Exempel 6.12

Betrakta försöket att kasta två tärningar i exempel 6.8. Variansen fås genom

x	$f(x)$	$x \cdot f(x)$	$x^2 \cdot f(x)$
2	1/36	2/36	4/36
3	2/36	6/36	18/36
4	3/36	12/36	48/36
5	4/36	20/36	100/36
6	5/36	30/36	180/36
7	6/36	42/36	294/36
8	5/36	40/36	320/36
9	4/36	36/36	324/36
10	3/36	30/36	300/36
11	2/36	22/36	242/36
12	1/36	12/36	144/36
Summa		252/36	1974/36

$$V(X) = \frac{1974}{36} - 7^2 = \frac{210}{36} = 5.83 \blacksquare$$

Övning 6.1

Låt X vara en stokastisk variabel som antar värdena 0 och 1 med sannolikheterna 0.3 respektive 0.7. Beräkna väntevärdet $E(X)$ och variansen $V(X)$.

Övning 6.2

Beräkna väntevärdet och variansen för den stokastiska variabeln X i exempel 6.3.

Övning 6.3

Vid ett försök inträffar antingen utfallet A eller utfallet B varvid $P(A) = 1/3$ och $P(B) = 2/3$. Försöket upprepas fyra gånger varvid upprepningarna sker oberoende av varandra. Beräkna sannolikheterna för följande händelser:

a) A inträffar minst en gång

b) A inträffar minst två gånger

c) Till en av upprepningarna av försöket definieras en stokastisk variabel X sådan att $X = 1$ om A inträffar och $X = 0$ om B inträffar. Bestäm sannolikhetsfördelningen för X och beräkna väntevärde och varians för X.

d) Definiera en stokastisk variabel Y som antal gånger utfallet A inträffar bland de fyra upprepningarna. Skriv upp frekvensfunktionen och fördelningsfunktionen för Y. Lös uppgifterna a och b med hjälp av sannolikhetsfördelningen för Y.

e) Bestäm väntevärde och varians för Y.

6.5 Några räkneregler för väntevärden och varianser

Ibland har man glädje av att transformera en stokastisk variabel till en annan genom ett matematiskt uttryck. Om t ex vinsten vid försäljning av en viss produkt är 50 kr per såld enhet och antalet sålda enheter är X, så blir den sammanlagda vinsten $50X$ kr. Om antalet sålda enheter är stokastiskt kan vi vara intresserade av att beräkna förväntad vinst, $E(50X)$ och variansen för vinsten $V(50X)$. Ett annat exempel är då R är en observation på radien i en cirkel. Vi får då cirkelns area A genom transformationen $A = \pi R^2$. Observera att om radien R betraktas som en stokastisk variabel, så är arean A också en stokastisk variabel. Det kan då vara av intresse att beräkna väntevärdet och variansen för cirkelns area. Det första exemplet är en linjär transformaton av en stokastisk variabel medan det andra exemplet är en olinjär transformation. I det här avsnittet skall vi ge väntevärden och varianser för linjära transformationer av stokastiska variabler.

1. Vänteväret av en konstant c är lika med konstanten själv:

$$E(c) = c.$$

2. Väntevärdet av produkten av en konstant c och en stokastisk variabel X är lika med produkten av konstanten och väntevärdet av den stokastiska variabeln:

$$E(cX) = cE(X).$$

3. Väntevärdet för summan av en konstant c och en stokastisk variabel X är lika med summan av konstanten och väntevärdet för den stokastiska variabeln:

$$E(c + X) = c + E(X).$$

4. Variansen för en konstant c är lika med noll:

$$V(c) = 0.$$

5. Variansen för produkten av en konstant c och en stokastisk variabel X är lika med produkten av konstanten i kvadrat och variansen av den stokastiska variabeln:

$$V(cX) = c^2 V(X).$$

6. Variansen för summan av en konstant c och en stokastisk variabel X är lika med variansen för den stokastiska variabeln:

$$V(c+X) = V(X).$$

Övning 6.4

Betrakta försöket att kasta två tärningar och låt X vara den stokastiska variabeln "pricksumman". I exemplen 6.10 och 6.12 har väntevärdet respektive variansen för X beräknats.

a) Bilda en ny stokastisk variabel, Y, genom

$$Y = X - E(X)$$

och beräkna väntevärdet och variansen av Y.

b) Bilda ytterligare ev ny stokastisk variabel, Z, genom

$$Z = \frac{Y}{\sqrt{V(Y)}}$$

och beräkna väntevärdet och variansen för Z.

Övning 6.5

Låt X vara en godtycklig stokastisk variabel med variansen $V(X) = \sigma^2 < \infty$. Visa att den stokastiska variabeln Z definierad genom

$$Z = \frac{X - E(X)}{\sigma}$$

har väntevärdet 0 och variansen 1.

Övning 6.6

Ett företag marknadsför två varor A och B. Sannolikhetsfördelningen för efterfrågan av varorna ges av nedanstående tabell. X och Y betecknar här antalet enheter av respektive vara.

Efterfrågan av vara A

x	0	1	2	3	4	5	6
$f(x)$	0.05	0.07	0.20	0.25	0.30	0.10	0.03

Efterfrågan av vara B							
y	0	1	2	3	4	5	6
$f(y)$	0.10	0.12	0.15	0.20	0.25	0.10	0.08

a) Beräkna sannolikheten att ingen vara efterfrågas en viss dag. Ange de antaganden du måste göra för beräkningarna.

b) Beräkna sannolikheten att minst tre enheter av A och högst fyra enheter av B efterfrågas en viss dag.

c) Beräkna sannolikheten att högst tre enheter av A och minst tre enheter av B efterfrågas en viss dag,

d) Beräkna sannolikheten att exakt tre enheter efterfrågas en viss dag och att dessa tre är av samma varuslag.

e) Vad är medelantalet efterfrågade enheter per dag för de två produkterna? Vad är varianserna? Illustrera sannolikhetsfördelningarna grafiskt och tolka varianserna.

f) Antag att företaget varje dag lagerför sex enheter av vara A och tio enheter av vara B och att vinsten vid varje såld vara är 100 kr, medan lagerkostnaden för osålda varor är 50 kr per enhet och dag. Beräkna förväntade vinsten för varje vara. Vilken vara är mer lönsam? Jämför detta med resultatet i uppgift e.

Övning 6.7

En anläggningsfirma överväger att åta sig ett arbete som ger 180.000 kr i vinst om arbetet blir färdigt i tid eller 60.000 kr i förlust om arbetet blir försenat.

a) Antag att sannolikheten att arbetet blir färdigt inom kontraktstiden är 0.8. Bör företaget åta sig arbetet?

b) Hur stor skall sannolikheten att arbetet blir färdigt inom kontraktstiden vara för att den förväntade vinsten skall bli lika med noll?

Övning 6.8

Sannolikheten att herr Johansson säljer sitt hus med en förlust av 5.000 kr är 3/18 och sannolikheten att han gör en vinst med 5.000 kr är 7/18 och en vinst med 10.000 kr är 3/18. Vad är Johanssons förväntade vinst vid husförsäljningen?

Övning 6.9

Ett försäkringsbolag skall försäkra en diamant värd 500.000 kr till sitt fulla värde. Premien är 4.000 kr per år. Antag att sannolikheten att diamanten blir stulen är 0,005 och låt X vara försäkringsbolagets årliga vinst.

a) Skriv upp sannolikhetsfördelningen för den stokastiska variabeln X.

b) Beräkna försäkringsbolagets förväntade vinst per år.

c) Bestäm den premie som ger försäkringsbolaget en förväntad årlig vinst på 10.000 kr.

Övning 6.10

Stig Gänglund är en hängiven spelare på hästar. Han befinner sig på en travbana och överväger att spela på en av tre hästar, Aldebaran, Blossom Star och Cassiopeja. Efter att ha studerat hästarna bedömer han sannolikheten att Aldebaran vinner till 0,2, att Blossom Star vinner till 0,4 och att Cassiopeja vinner till 0,1. Om Aldebaran vinner ger det en vinst på 2,5 gånger insatsen, om Blossom Star vinner ger det 1,2 gånger instasen och om Cassiopeja vinner ger det 4 gånger insatsen.

a) Beräkna förväntad vinst vid spel på var och en av de tre hästarna om Stig satsar 100 kr.

b) Bör Stig satsa på någon av hästarna och i så fall vilken?

6.6 Simultant fördelade stokastiska varabler

I många tillämpningar förekommer modeller med flera stokastiska variabler och syftet med många analyser är att studera eventuella samband mellan två eller flera variabler. Ur ekonomisk teori följer att efterfrågan av en viss vara (eller tjänst) beror av konsumenternas inkomst och av varans pris. Konsumtionen av en viss vara är förmodligen olika för olika ålderskategorier av konsumenter. Värdet av aktier i olkia företag kan antingen ha en tendens att öka samtidigt eller att värdet av någon aktie ökar då värdet av andra aktier minskar osv. Kunskap om hur olika aktiers värden samvarierar är av stor betydelse då aktieportföljer skall konstrueras. I detta avsnitt diskuteras grunderna för sannolikhetsmodeller för två stokastiska variabler.

Låt X och Y vara två diskreta stokastiska variabler. Då definieras den *simultana frekvensfunktionen*, $f_{X,Y}(x, y)$, av

$$f_{X,Y}(x, y) = P(X = x \cap Y = y).$$

Vi skriver X, Y som index till f för att tydliggöra att frekvensfunktionen f avser den simultana fördelningen för X och Y.

Exempel 6.13

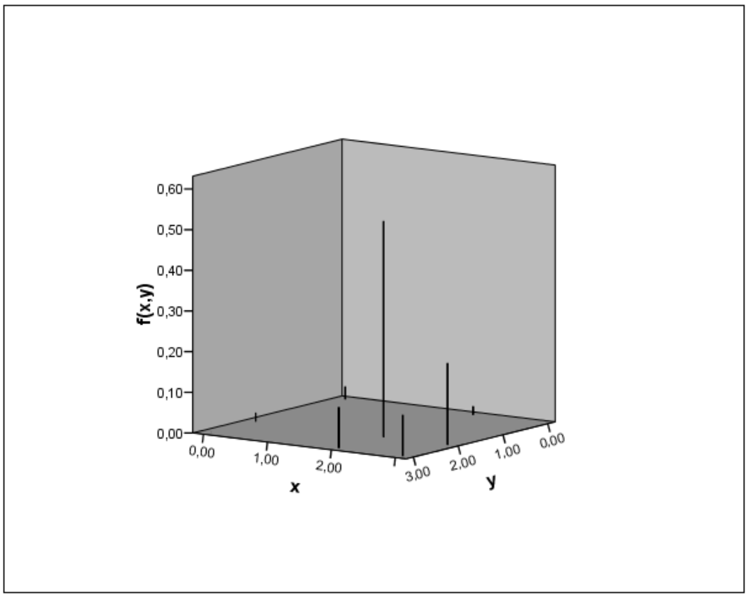
Vid fastighetstaxering av småhus värderas husen och ett taxeringsvärde fastställs. Taxeringsvärdet skall vara c:a 75% av marknadsvärdet. Taxeringsvärdet bestäms utifrån ett poängsystem där mer välutrustade hus får högre poäng. Om t ex huset är utrustat med värmepump får huset 3 poäng, om det har konventionell uppvärmning får det 2 poäng och om det saknar uppvärmning får det 0 poäng. Vidare, om huset har isolerglas i fönstren får huset 3 poäng, om det har två eller treglasfönster får det 2 poäng och om det har enkelglas i fönstren får det 0 poäng. Tabellen nedan visar andelen småhus i ett visst område fördelade över olika kategorier av värmesystem och fönsterkvaliteter. Vi kan här tänka oss att X är den poäng ett slumpmässigt valt småhus från området får för sitt värmesystem och Y den poäng huset får för den sorts fönster huset har. Tabellen visar t ex att sannolkheten att $X = 3$ och $Y = 2$ (ett slumpmässigt valt småhus har värmepump och är utrustat med två eller treglasfönster) är

$$f_{X,Y}(3,2) = P(X = 3 \cap Y = 2) = 0,20 \qquad \blacksquare$$

Tabell 6.2 Simultana sannolikhetsfördelningen för poäng för värmesystem (X) och typ av fönster (Y)

			X			Marginal-
			Saknar uppv	Konv uppv	Värmepump	fördeln för
			0	2	3	Y
Y	Enkelglas	0	0,03	0,02	0,00	0,05
	Två el treglas	2	0,02	0,53	0,20	0,75
	Isolerglas	3	0,00	0,10	0,10	0,20
Marginalfördeln för X			0,05	0,65	0,30	1,00

Grafiskt kan den simultana frekvensfunktionen illustreras med ett tvådimensionellt stolpdiagram. För fördelningen av poäng för utrustning av småhus i exempel 6.13 kan den simultana frekvensfunktionen illustreras som i figur 6.7.



Figur 6.7 Simultana frekvensfunktionen från exempel 6.13

Notera att värdet hos en simultan frekvensfunktion måste vara större än eller lika med noll eftersom dess värden är sannolikheter,

$$f_{X,Y}(x, y) \geq 0, \text{ för alla } x, y$$

Vidare måste summan av alla värden hos en simultan frekvensfunktion vara lika med ett eftersom vi då summerar sannolikheterna för alla utfall i ett utfallsrum,

$$\sum_x \sum_y f(x, y) = 1$$

där $\sum_x \sum_y$ betyder att vi summerar över alla värden som X kan anta och över alla värden som Y kan anta.

För kontinuerliga stokastiska variabler definieras den simultana täthetsfunktionen på motsvarande sätt, men här måste vi komma ihåg att sannolikheter svarar mot volymer som definieras av den tvådimensionella täthetsfunktionen. För två intervall A och B gäller att

$$P(X \in A \cap Y \in B) = \int_B \int_A f_{X,Y}(x, y)\, dx dy$$

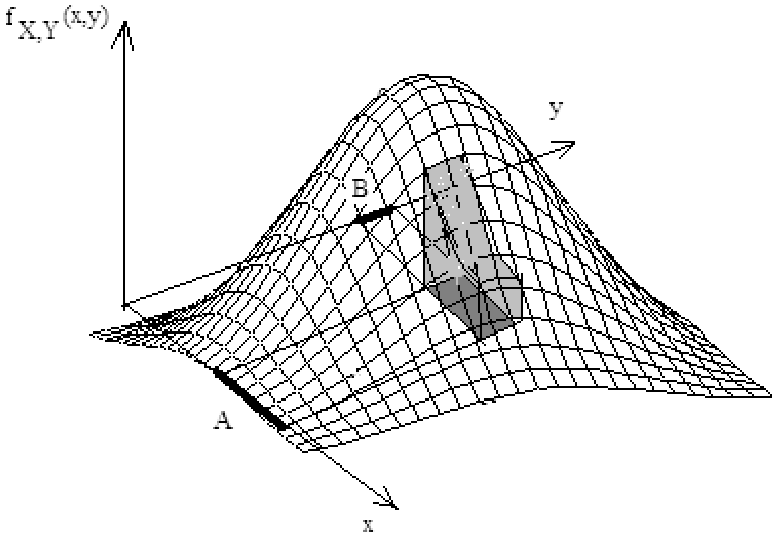
se figur 6.8.

För simultana täthetsfunktioner gäller det att deras värden får inte vara negativa

$$f_{X,Y}(x, y) \geq 0, \text{ för alla } x, y$$

annars skulle vi kunna få negativa sannolikheter, vilket strider mot Kolmogorovs axiom. Vidare måste integralen av täthetsfunktionen över hela utfallsrummet vara 1,

$$\int_y \int_x f_{X,Y}(x, y)\, dx dy = 1.$$



Figur 6.8 Simultana täthetsfunktionen för två kontinuerliga stokastiska variabler. Sannolikheten att X hamnar i intervallet$A$ och Y hamnat i intervallet B svarar mot en vlym under täthetsfunktionen.

Detta kommer sig av att sannolikheten att X får ett värde i utfallsrummet och att Y får ett värde i utfallsrummet är en säker händelse, Ω, och skall därför ha sannolikheten 1.

På samma sätt som vi definierar marginalsannolikheter för händelser i föregående kapitel definierar vi marginalsannolikheter för stokastiska variabler. Marginalfördelningen för en stokastisk variabel får vi genom att summera sanolikheterna respektive integrera tätheterna för ett visst värde på variabeln. För diskreta variabler får vi alltså marginalfördelningen för X genom

$$f_X(x) = \sum_y f_{X,Y}(x,y)$$

där $\sum\limits_y$ innebär att vi summerar över alla värden som variabeln Y kan anta.

För kontinuerliga stokastiska variabler får vi analogt

$$f_X(x) = \int_y f_{X,Y}(x,y)\,dy$$

Exempel 6.13 (fortsättning)

Marginalfördelningarna för X och Y anges i marginalerna i tabell 6.2. Notera att vi får marginalerna genom att summera sannolikheterna radvis respektive kolumnvis. Notera också att marginalerna summerar till 1. Sannolikheterna av alla utfall i ett utfallsrum skall ju summera till 1. ■

Den betingade fördelningen för en diskret stokastisk variabel, givet ett visst värde på en annan diskret stokastisk variabel är samlingen av betingade sannolikheter

$$f_{Y|X}(y \mid x) = P(Y = y \mid X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Här använder vi $X \mid Y$ som index på frekvensfunktionen för att markera att den anger betingade sannolikheter för Y giver X.

Exempel 6.13 (fortsättning)

Bland de småhus som har värmepump installerad, $X = 3$, är det $f_{Y|X}(2,3) = f_{X,Y}(3,2)/f_X(3) = 0,20/0,30 = \frac{2}{3} \approx 0,67$ som har två eller treglasfönster, $Y = 2$.

För varje värde på X kan vi definiera en betingad sannolikhetsfördelning för Y, de poäng småhusen får för den typ av fönster som är installerade. De betingade fördelningarna givet $X = 0$, ingen uppvärmning, $X = 2$, konventionell uppvärmning och $X = 3$, värmepump, visas i tabell 6.3. Observera att varje betingad sannolikhet summerar till 1. ■

Tabell 6.3 Betingade sannolikhetsfördelningar för Y givet X i exempel 6.13

		Saknar uppvärmn $X=0$	Konv uppvärmn $X=2$	Värmepump $X=3$
Enkelglas	$Y=0$	0,60	0,03	0,00
Två el treglas	$Y=2$	0,40	0,82	0,67
Isolerglas	$Y=3$	0,00	0,15	0,33
Summa		1,00	1,00	1,00

Om vi jämför de betingade fördelningarna för Y givet X i exempel 6.13 ser vi att de är olika för olika värden på X. Det t ex mycket vanligare att hus har enkelglas i fönstren då de saknar uppvärmning än då de har konventionell uppvärmning eller värmepump. Det faktum att vi får olika betingade fördelningar för Y beroende på vilket värde på X vi betingar på, tar vi som definition på att X och Y är stokastiskt beroende. Om däremot de betingade fördelningarna för Y vore lika oavsett vilket värde på X vi betingar på, vore X och Y stokastiskt oberoende.. Vi säger alltså att X och Y är stokastiskt oberoende om

$$f_{Y|X}(y \mid x) = f_Y(y)\,, \text{ för alla } x$$

Observera att eftersom

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

följer att X och Y är stokastiskt oberoende om

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)\,, \text{ för alla } x \text{ och } y.$$

För kontinuelliga stokastska variabler definierar vi den betingade täthetsfunktionen analogt genom

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

och två kontinuerliga stokastiska variabler är stokastiskt beroende om den betingade tätheten för den ena variabeln, t ex Y, givet den andra variabeln, t ex X, är olika för olika värden på X. Variablerna är stokastiskt oberoende om

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)\,, \text{ för alla } x \text{ och } y.$$

Betingade väntevärden definieras som väntevärden i betingade fördelningar. Sålunda definieras väntevärdet för Y givet $X = x$ som väntevärdet i den betingade fördelningen för Y givet $X = x$. Om X och Y är diskreta stokastiska variabler innebär detta att

$$E(Y \mid X = x) = \sum_y y f_{Y|X}(y \mid x)$$

Exempel 6.13

Om $X = 0$ kan vi beräkna det betingade väntevärdet för Y enligt tabellen nedan.

y	$f_{Y\|X}(y \mid 0)$	$y f_{Y\|X}(y \mid 0)$	$y^2 f_{Y\|X}(y \mid 0)$
0	0,60	0,00	0,00
2	0,40	0,80	1,60
3	0,00	0,00	0,00
Summa	1,00	0,80	1,60

Tygligen är $E(Y \mid X = 0) = 0,80$ dvs medelpoängen för fönsterkvalitet är 0,80 bland de som saknar uppvärmning, $X = 0$. På motsvarande sätt kan vi beräkna de betingade väntevärdena givet $X = 2$ konventionell uppvärmning, och $X = 3$, värmepump. Resultaten visas i tabell 6.4. Tydligen ökar det betingade väntevärdet för Y då värdet på X ökar. ■

Tabell 6.4 Betingade väntevärden för Y givet X i exempel 6.13

	X		
	Saknar uppv 0	Konvent uppv 2	Värmepump 3
$E(Y \mid X = x)$	0,80	2,09	2,33

Om vi viktar ihop de betingade väntevärdena för Y givet X med sannolikheterna för de olika värdena för X (margnalfördelningen för X) får vi i exempel 6.13

$$0,80 \cdot 0,05 + 2,09 \cdot 0,65 + 2,33 \cdot 0,30 = 2,10$$

Detta visar sig nu vara detsamma som förväntat värde i marginalfördelningen för Y. Allmänt gäller för diskreta stokastiska variabler att

$$\begin{aligned} E\left(E\left(Y \mid X=x\right)\right) &= \sum_x E\left(Y \mid X=x\right) f_X\left(x\right) = \sum_x \left\{\sum_y y f_{Y|X}\left(y \mid x\right)\right\} f_X\left(x\right) \\ &= \sum_x \sum_y y \frac{f_{X,Y}\left(x,y\right)}{f_X\left(x\right)} f_X\left(x\right) = \sum_x \sum_y y f_{X,Y}\left(x,y\right) \\ &= \sum_y y \sum_x f_{X,Y}\left(x,y\right) = \sum_y y f_Y\left(y\right) = E\left(Y\right) \end{aligned}$$

Motsvarande regel gäller givetvis för kontinuerliga stokastska variabler, men summeringarna i beviset måste bytas ut mot integreringar.

6.7 Kovarians och korrelation

Kovarians är ett mått som används för att beskriva hur två stokastiska variabler samvarierar. Kovariansen spelar dessutom en viktig roll när man skall bestämma variansen för en linjär kombination, t ex summan, av två stokastiska variabler. I detta avsnitt definierar vi kovarians och därefter ett standardiserat sambandsmått som kallas korrelation.

Antag att X och Y är stokastiska variabler och betrakta produkten $C = (X - E(X))(Y - E(Y))$. Om det finns en stor sannolikhet att både X och Y samtidigt är större än sina respektive väntevärden blir C stor med stor sannolikhet. På samma sätt blir C stor med stor sannolikhet om både X och Y samtidigt är mindre än sina väntevärden med stor sannolikhet. Förväntat värde av C kommer då att bli stort. Om, å andra sidan, det finns en stor sannolikhet att X är större än sitt väntevärde samtdigt som Y är mindre än sitt väntevärde, blir C stort negativt med stor sannolikhet. Likaså blir C stor negativt med stor sannolikhet om X är mindre än sitt väntevärde samtidigt som Y är större än sitt väntevärde med stor sannolikhet. I det fallet kommer väntevärdet av C att bli stort negativt. Allmänt definierar vi väntevärdet av C som kovariansen mellan X och Y. För diskreta stokastiska variabler får vi då

$$Cov\,(X, Y) = \sum_x \sum_y (x - E\,(X))\,(y - E\,(Y))\,f_{X,Y}\,(x, y)$$

och för kontinuerliga stokastiska variabler

$$Cov\,(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E\,(X))\,(y - E\,(Y))\,f_{X,Y}\,(x, y)\,dxdy$$

Ibland är det enklare att beräkna kovariansen med hjälp av följande formel

$$Cov\,(X, Y) = E\,(XY) - E\,(X)\,E\,(Y)$$

Exempel 6.13 (fortsättning)

För att beräkna kovariansen mellan variablerna X, poäng för typ uppvärmning, och Y, poäng för typ av fönster, i exemplet med beräkning av taxeringsvärde av småhus, kan vi först konstruera en tabell över produkten av poäng, xy, se nedanstående tabell.

		X		
		0	2	3
	0	0	0	0
Y	2	0	4	6
	3	0	6	9

Nästa steg är att multiplicera talen i tabellen med motsvarande sannolikheter redovisade i tabell 6.2. Detta ger

		X		
		0	2	3
Y	0	0,00	0,00	0,00
	2	0,00	2,12	1,20
	3	0,00	0,60	0,90

Summerar vi nu alla tal i denna tabell får vi $E(XY)$, dvs vi får

$$E(XY) = 2,12 + 1,20 + 0,60 + 0,90 = 4,82$$

så att

$$Cov(X, Y) = 4,82 - 2,20 \cdot 2,10 = 0,20 \blacksquare$$

Om $Cov(X, Y)$ är stor innebär det att X och Y tenderar att vara stora samtidigt och att vara små samtidigt. Vi säger då att vi har ett positivt samband mellan X och Y. På motsvarande sätt innebär ett stort negativt värde på $Cov(X, Y)$ att X tenderar att vara stor då Y är liten och vice versa. Vi säger då att vi har ett negativt samband mellan X och Y. Man kan visa att om X och Y är stokastiskt oberoende så är $Cov(X, Y) = 0$. Omvändningen, dvs att dra några slutsatser om eventuellt oberoende mellan X och Y då $Cov(X, Y) = 0$ följer inte helt automatiskt, vilket följande enkla teoretiska exempel visar.

Exempel 6.14

Antag att X är en likformigt fördelad stokastisk variabel som kan anta värdena -1, 0 och 1. Låt vidare $Y = X^2$, dvs det finns ett exakt matematiskt samband mellan X och Y, vet vi värdet på X så vet vi även värdet på Y. Vi får att

$$E(X) = \frac{1}{3} \cdot (-1) + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 = 0$$

$$E(Y) = \frac{1}{3} \cdot (-1)^2 + \frac{1}{3} \cdot 0^2 + \frac{1}{3} \cdot 1^2 = \frac{2}{3}$$

$$E(XY) = \frac{1}{3} \cdot (-1) \cdot (-1)^2 + \frac{1}{3} \cdot 0 \cdot 0 + \frac{1}{3} \cdot 1 \cdot 1 = 0$$

$$Cov\,(X, Y) = E(XY) - E(X)\,E(Y) = 0 - 0 \cdot \frac{2}{3} = 0$$

dvs kovariansen är 0 trots att det råder ett starkt (olinjärt) samband mellan X och Y. ■

Kovariansen kan många gånger vara svår att tolka som mått på samvariation. Kovariansens storlek beror nämligen inte bara på styrkan av ett linjärt samband mellan variablerna utan också på de ingående variablernas varianser. Man brukar därför definiera en standardiserad variant av kovariansen, kallad korrelation. Korrelationen mellan två stokastiska variabler definieras genom

$$\rho = \frac{Cov\,(X, Y)}{\sqrt{V(X)\,V(Y)}}$$

dvs kovariansen dividerad med variablernas standardavvikelser.

Exempel 6.13 (fortsättning)

Vi har tidigare räknat ut att $Cov\,(X, Y) = 0,20$, $V(X) = 0,46$ och $V(Y) = 0,39$ så att

$$\rho = \frac{0,20}{\sqrt{0,46 \cdot 0,39}} = 0,47 \ ■$$

Korrelationen är standardiserad på så sätt att dess värden ligger i intervallet $[-1.1]$. En korrelation nära 1 indikerar ett starkt positivt samband och en korrelation nära -1 indikerar ett starkt negativt samband. Liksom för kovariansen gäller att om X och Y är oberoende är korrelationen noll.

En linjär kombination av två stokastiska variabler X och Y är uttryck av formen

$$aX + bY$$

där a och b är konstanter. Som ett exempel får vi summan av två variabler om vi väljer $a = b = 1$ och vi får skillnaden om vi väljer $a = 1$ och $b = -1$. Viktiga räkneregler för väntevärden och varianser av linjära kombinationer är

$$E(aX + bY) = aE(X) + bE(Y)$$
$$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abCov(X, Y).$$

Väntevärdet för en linjär kombination av stokastiska variabler är helt enkelt samma linjära kombination av väntevärden. Uttrycket för variansen av en linjär kombination är dock lite mer komplicerad. Dels ser vi att konstanterna a och b kvadreras och multipliceras med varianserna. Detta är i enlighet med uttrycket för variansen för en konstant multiplicerat med en stokastisk variabel som vi diskuterade i avsntt 6.5. Dels ser vi att variansen också beror på kovariansen. Om X och Y är stokastiskt oberoende är $Cov(X, Y) = 0$ och formeln reduceras till

$$V(aX + bY) = a^2 V(X) + b^2 V(Y).$$

Exempel 6.13 (fortsättning)

Den sammanlagda poäng småhus får vid fastghetstaxeringen för det värmesystem och den typ av fönster som finns i huset har väntevärdet

$$E(X + Y) = 2,20 + 2,10 = 4,30$$

och variansen

$$V(X + Y) = 0,46 + 0,39 + 2 \cdot 0,20 = 1,25$$

Övning 6.11

Betrakta följande simultana sannolikhetsfördelning

		X	
		0	1
Y	0	0,30	0,20
	1	0,20	0,30

a) Beräkna marginalfördelningarna för X och Y

b) Beräkna betingade fördelningen för Y givet X

c) Beräkna betingade väntevärdet för Y givet X

d) Avgör om X och Y är stokastiskt oberoende

e) Bestäm kovariansen och korrelationen

f) Bilda en ny variabel, $Z = X + Y$, och bestäm sannolikhetsfördelningen för Z

g) Beräkna väntevärde och varians för Z dels med hjälp av sannolikhetsfördelningen för Z och dels med hjälp av räknereglerna för väntevärde och varians för linjära kombinationer av stokastiska variabler

Övning 6.12

Företaget "Mutter och skruv" tillverkar och säljer två produkter: muttrar och skruvar. Företagets vinst vid försäljning av ett pari muttrar är 3000 kr och vid försäljning av ett parti skruvar är vinsten 5000 kr. Företagets VD, Stig Gänglund, har satt upp följande sannolikheter för försäljningen under en vecka:

		Antal sålda partier muttrar				
		0	1	2	3	4 eller fler
Antal	0	0,05	0,10	0,05	0,00	0,00
sålda	1	0,05	0,15	0,04	0,01	0,00
partier	2	0,00	0,10	0,20	0,05	0,00
skruvar	3	0,00	0,05	0,05	0,10	0,00
	4 eller fler	0,00	0,00	0,00	0,00	0,00

a) Beräkna sannolikheten att den sammanlagda försäljningen av muttrar och skruvar en vecka blir tre partier

b) Avgör om försäljningen av skruvar och muttrar är oberoende av varandra.

c) Beräkna sannolikheten att vinsten av försäljningen av muttrar blir större än vinsten av försäljningen av skruvar

d) Beräkna förväntad vinst från den sammanlagda försäljningen av muttrar och skruvar

Övning 6.13

Två investeringar, A och B, kan antingen resultera i en förlust om 1000 kr eller en vinst om 2000 kr vardera. Sannolikheten att investering A ger vinst är 0,4 och att B ger vinst är 0,6 och att båda ger vinst är 0,15. Låt X vara vinsten från investering A och Y vara vinsten från investering B.

a) Bestäm simultana sannolikhetsfördelningen för X och Y

b) Beräkna kovarians och korrelation

c) Beräkna väntevärde och varians för summan $X + Y$

6.8 Tillämpning: Analys av värdepappersportföljer

Det teori vi har gått igenom i detta kapitel kan bl a användas för att sätta samman aktier och andra värdepapper till en s k värdepappersportfölj och för att analysera innehållet i en värdepappersportfölj. Idén med att kombinera flera värdepapper är att man då kan konstruera en investering med en viss bestämd förväntad avkastning och en viss risk. En portfölj med t ex aktier som förväntas öka i värde om andra aktier i portföljen minskar i värde, ger en portfölj med tämligen stabilt värde. En nedgång i en aktie kompenseras av en uppgång i en annan. Portföljens värde kommer troligen inte att göra några stora språng, varken uppåt eller neråt. Man säger då att man har en portfölj med liten risk. Om däremot alla aktier i en portfölj antingen väntas öka samtidigt i värde eller minska samtdigt i värde, kommer förändringarna i de olika aktiernas värden att förstärka varandra. Vi får då en portfölj vars värde kan öka kraftigt eller minska kraftigt. En sådan portfölj sägs ha en stor risk.

För att förenkla analysen av värdepappersportföljer tänker vi oss här en portfölj med endast två aktier, A och B. Vi tänker också att avkastningen från aktierna kan beskrivas med två diskreta stokastiska variabler, X respektive Y. Den simultana sannolikhetsfördelningen för X och Y framgår av tabellen nedan.

		x		
		500 kr	700 kr	900 kr
	500 kr	0,15	0,10	0,05
y	700 kr	0,10	0,20	0,10
	900 kr	0,05	0,10	0,15

Marginalfördelningarna för X och Y är lika och framgår i nedanstående tabell.

x	$f_X(x)$	$xf_X(x)$	$x^2 f_X(x)$
500	0,30	150	75000
700	0,40	280	196000
900	0,30	270	243000
Summa	1,00	700	514000

y	$f_Y(y)$	$yf_Y(y)$	$y^2 f_Y(y)$
500	0,30	150	75000
700	0,40	280	196000
900	0,30	270	243000
Summa	1,00	700	514000

Tydligen är $E(X) = E(Y) = 700$ och $V(X) = V(Y) = 514000 - 700^2 = 24000$. För att beräkna kovariansen konstruerar vi först en tabell med produkterna $xyf_{X,Y}(x, y)$ och summerar sedan alla cellerna för att få $E(XY) = 498000$.

		x		
		500 kr	700 kr	900 kr
	500 kr	37500	35000	22500
y	700 kr	35000	98000	63000
	900 kr	22500	63000	121500

Detta ger nu $Cov(X, Y) = 498000 - 700 \cdot 700 = 8000$ och korrelationen $\rho = 8000/\sqrt{24000 \cdot 24000} = 0,33$.

Antag nu att vi har en portfölj bestående av 10 enheter av aktie A och 15 enheter av aktie B. Portföljens avkastning är därmed

$$Z = 10X + 15Y$$

den förväntade avkastningen blir

$$E(Z) = 10E(X) + 15E(Y) = 10 \cdot 700 + 15 \cdot 700 = 17500$$

och risken, mätt i varians blir

$$\begin{aligned} V(Z) &= 10^2 V(X) + 15^2 V(Y) + 2 \cdot 10 \cdot 15 \cdot Cov(X, Y) \\ &= 100 \cdot 24000 + 225 \cdot 24000 + 300 \cdot 8000 = 10200000 \end{aligned}$$

Standardavvikelsen av Z är 3194 kr. Antag nu att vi har en annan protfölj som även den består av 10 enheter av aktie A men i stället för 15 enheter av aktie B finns det 15 enheter av en aktie C. Beteckna nu avkastningen från aktie C med U och antag att följande simultana fördelning gäller

			x	
		500 kr	700 kr	900 kr
	500 kr	0,05	0,10	0,15
u	700 kr	0,10	0,20	0,10
	900 kr	0,15	0,10	0,05

Här ser vi att marginalfördelningen för U blir

u	$f_U(u)$	$uf_U(u)$	$u^2 f_U(u)$
500	0,30	150	75000
700	0,40	280	196000
900	0,30	270	243000
Summa	1,00	700	514000

dvs densamma som marginalfördelningarna för X och Y, men kovariansen mellan X och U blir annorlunda:

		x		
		500 kr	700 kr	900 kr
	500 kr	37500	35000	22500
u	700 kr	35000	98000	63000
	900 kr	22500	63000	121500

Summerar vi alla cellerna i den tabellen får vi $E(XU) = 481500$ och kovariansen blir $Cov(X, U) = 481500 - 700 \cdot 700 = -8500$. Detta ger en korrelation mellan X och U som är $\rho = -8500/\sqrt{24000 \cdot 24000} = -0,35$. Vi bildar nu en ny portfölj genom

$$T = 10X + 15U$$

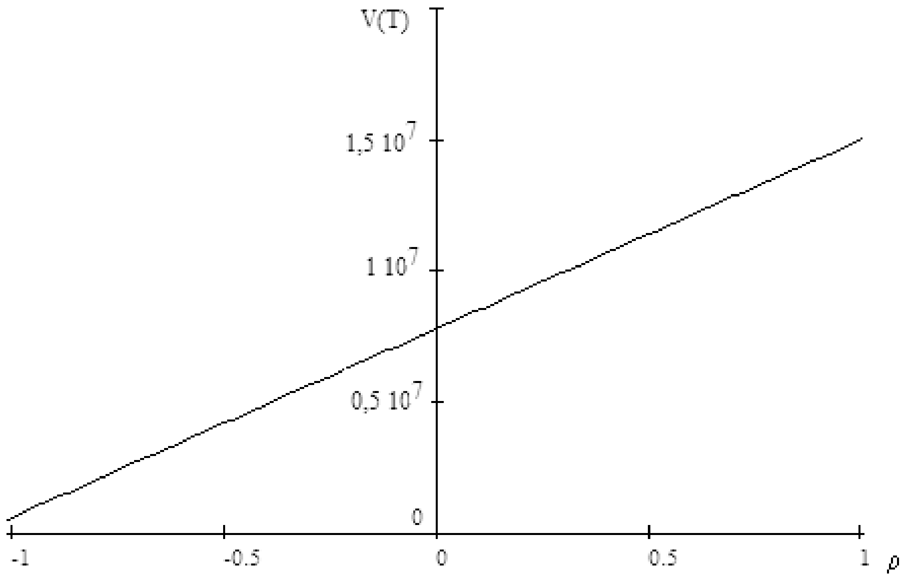
med den förväntade avkastningen

$$E(T) = 10E(X) + 15E(U) = 10 \cdot 700 + 15 \cdot 700 = 17500$$

dvs samma som för den tidigare portföljen, men risken, mätt i varians blir

$$\begin{aligned} V(T) &= 10^2 V(X) + 15^2 V(U) + 2 \cdot 10 \cdot 15 \cdot Cov(X, U) \\ &= 100 \cdot 24000 + 225 \cdot 24000 + 300 \cdot (-8500) = 5250000 \end{aligned}$$

Standardavvikelsen har nu sjunkit till ungefär 2291 kr. Risken är tyligen beroende av hur stor korrelationen är mellan de två aktierna som ingår i portföljen. I figur 6.9 visas hur risken beror av korrelationen. Där framgår att risken är minst om korrelationen är -1, dvs förändringar i aktiernas värde motverkar varandra exakt så att när den ena ökar sitt värde minskar den andra och vice versa.



Figur 6.9 Variansen för avkastningen som funktion av korrelationen

Även antalet aktier av vardera slaget påverkar risken, vilket följande exempel visar. Antag att vi sätter samman en protfölj bestående av 5 enheter av aktie A och 20 enheter av aktie C, så att avkastningen av den portföljen blir

$$S = 5X + 20U$$

Förväntade avkastningen blir då

$$E(S) = 5E(X) + 20E(U) = 5 \cdot 700 + 20 \cdot 700 = 17500$$

dvs samma som för de tidigare två portföljerna, men risken, mätt i varians blir

$$\begin{aligned} V(S) &= 5^2 V(X) + 20^2 V(U) + 2 \cdot 5 \cdot 20 \cdot Cov(X, U) \\ &= 25 \cdot 24000 + 400 \cdot 24000 + 100 \cdot (-8500) = 8500000 \end{aligned}$$

Vi ser att risken har nu ökat och standardavvikelsen är 2915. För att bestämma vilken kombination av aktierna A och C som ger den minsta risken kan vi tänka oss en protfölj med a stycken enheter av aktie A. Om totala antalet enheter fortfarande skall vara 25 blir det därmed $25 - a$ enheter av aktie B och portföljens avkastning blir

$$R = aX + (25 - a)\,U.$$

Förväntad avkastning blir samma som tidigare

$$\begin{aligned} E\left(R\right) &= aE\left(X\right) + (25 - a)\,E\left(U\right) = a \cdot 700 + (25 - a) \cdot 700 \\ &= 25 \cdot 700 = 17500 \end{aligned}$$

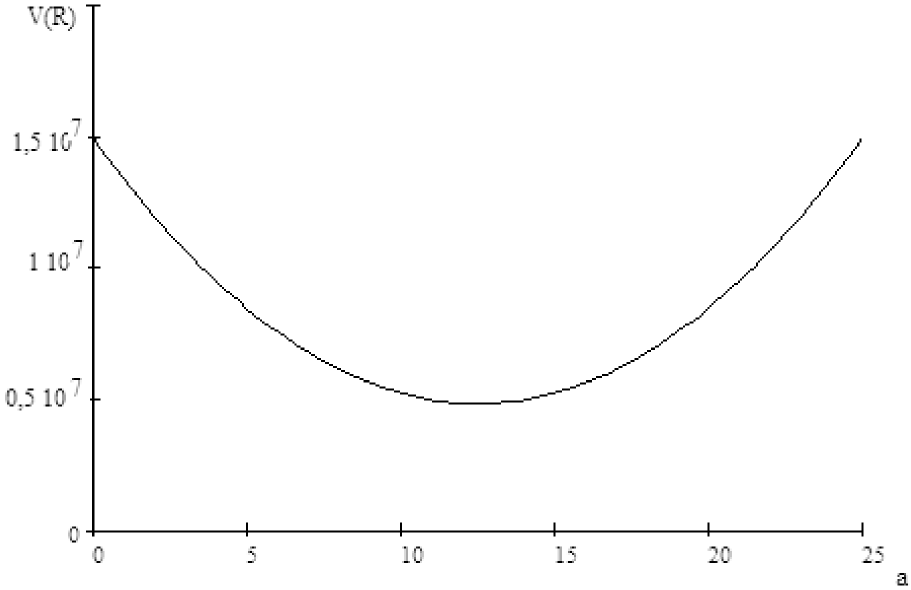
och risken blir

$$\begin{aligned} V\left(R\right) &= a^2 V\left(X\right) + (25 - a)^2\,V\left(U\right) + 2a\,(25 - a)\,Cov\left(X, U\right) \\ &= a^2 24000 + (25 - a)^2\,24000 + 2a\,(25 - a)\,(-8500) \\ &= 65000a^2 - 1625000a + 15000000 \end{aligned}$$

Detta är en andragradsfunktion vars graf visas i figur 6.10. För att finna det antal enheter som minimerar risken deriverar vi funktionen och sätter derivatan till noll. Detta ger ekvationen

$$130000a - 1625000 = 0$$

som har lösningen $a = 12,5$, dvs vi får i detta fall den minsta risken om vi tar lika många av varje aktie i portföljen. Detta resultat beror givetvis av hur stora varianserna för avkastningarna av de båda aktierna är.



Figur 6.10 Variansen för avkastningen som funktion av sammansättningen av akter