

Lösningförslag till uppgifter i
SANNOLIKHETSTEORI och STATISTIKTEORI med TILLÄMPNINGAR
av Blom, Enger, Englund, Grandell & Holst.

Version 18 december 2006

Fel i lösningarna mottages tacksamt till mattsson@math.kth.se.

Notera att lösningarna på vissa ställen utnyttjar andra, mer fullständiga, tabeller än vad som normalt är tillgängliga för studenterna. Därför kan t.ex. kvantiler i normalfördelningen och $t(n)$ -fördelningar i lösningarna vara bestämda med mycket god noggrannhet.

2.1

a) Notera att om tärningen har n sidor så måste vid $n + 1$ gjorda försök någon sida förekomma minst två gånger (Dirichlets lådprincip). Alltså, ett lämpligt utfallsrum är $\Omega = \{2, 3, 4, 5, 6, 7\}$ om tärningen har 6 sidor.

b) Vid varje försök (kast) finns möjligheten att man inte erhåller samma antal ögon som i föregående kast vilket gör att antalet nödvändiga kast är obegränsat och ett lämpligt utfallsrum är $\Omega = \{2, 3, 4, 5, \dots\}$.

2.2 Om den överskjutande längden efter en tillsågning är åtminstone 100 cm kommer även den att tillsågas för att skapa en ny plank. Detta upprepas till den kvarvarande längden är för kort att utgöra en meterlång plank.

Alltså, den överskjutande bitens längd kan sägas vara ett utfall i utfallsrummet $\Omega = \{x : 0 \leq x < 100 \text{ cm}\} = [0, 100)$.

2.3

a) Utfallsrummet består av 8 element.

$$\Omega = \{DDD, DDK, DKD, DKK, KDD, KDK, KKD, KKK\}$$

där $A = \{\text{exakt två defekta}\}$ består av utfallen $A = \{DDK, DKD, KDD\}$.

b) Utfallsrummet består av 4 element,

$$\Omega = \{0, 1, 2, 3\},$$

där $A = \{\text{exakt två defekta}\}$ består av utfallet $A = \{2\}$.

c) (1) Utfallsrummet är överuppräknligt och ges av

$$\Omega = \{x : x \geq 0\} = \mathbb{R}_+ = [0, \infty).$$

Händelsen är $\{x : a < x < b\} = (a, b)$.

(2) Utfallsrummet är överuppräknligt och kan ges av

$$\Omega = \{(x, y) : x \geq 0, y \geq 0\} = \mathbb{R}_+^2.$$

Händelsen är $\{(x, y) : x > a, y > a\} = (a, \infty) \times (a, \infty)$.

(3) Utfallsrummet är överuppräknligt och ges av

$$\Omega = \{(x_1, \dots, x_n) : x_1 \geq 0, \dots, x_n \geq 0\} = \mathbb{R}_+^n.$$

2.4 Händelsen $A \cap B$ består av utfallen som finns i A och i B , det vill säga händelsen kan uttryckas som "både A och B inträffar".

Händelsen $A \cap B^*$ består av utfallen i A men inte i B , det vill säga $A \cap B^*$ är händelsen "A men inte B" eller "endast A".

Händelsen $A^* \cap B^*$ består av de utfall som inte ligger i A och inte heller i B så händelsen beskrivs av "inte A och inte B ", det vill säga "varken A eller B ". (Notera att meningen "inte någon av A eller B " låter sig skrivas som $(A \cup B)^*$ vilket enligt de Morgan är ekvivalent med $A^* \cap B^*$.)

2.5 Utfallsrummet består av de 36 utfallen $\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\}$. Med beteckningarna

a) $A = \{\text{Poängsumma mindre än 6}\}$ så är $A = \{(x, y) \in \Omega : x + y < 6\}$, det vill säga A består av de 10 utfallen $\{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (4, 1)\}$. Således är $P(A) = 10/36$.

b) $B = \{\text{Samma poäng vid båda kasten}\}$ så är $B = \{(x, x) \in \Omega\}$, det vill säga B består av de 6 utfallen $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$. Således $P(B) = 1/6$.

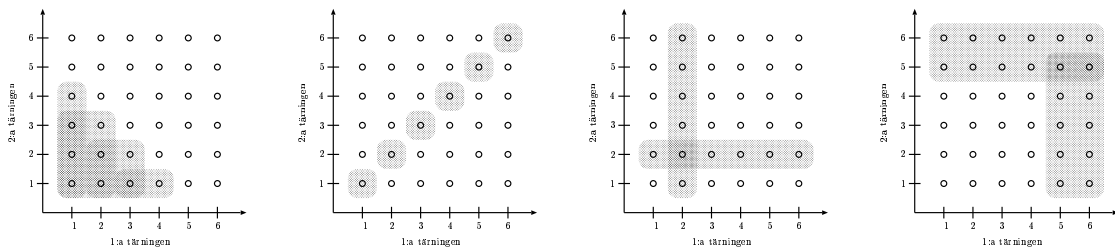
Man kan också tänka sig att det första kastet bestämmer det värde som man skall träffa i det andra kastet. För varje utfall av det första kastet är sannolikheten $1/6$ att kast 2 kommer att ge samma värde (kastet är oberoende).

c)

$$\begin{aligned} C &= \{\text{Åtminstone ett av kasten ger precis två poäng}\} \\ &= \{(1, 2), (2, 2), (3, 2), (4, 2), (5, 2), (6, 2), (2, 1), (2, 3), (2, 4), (2, 5), (2, 6)\} \end{aligned}$$

och $P(C) = 11/36$.

d) Slutligen, $D = \{\text{Åtminstone ett av kasten ger minst fem poäng}\}$ innehåller 20 distinkta utfall och $P(D) = 20/36$.



2.6 Enligt Kolmogorovs axiom är $P(A \cup B) = P(A) + P(B)$ då A och B är disjunkta, det vill säga $A \cap B = \emptyset$. Alltså är $P(B) = P(A \cup B) - P(A) = 0.75 - 0.25 = 0.50$.

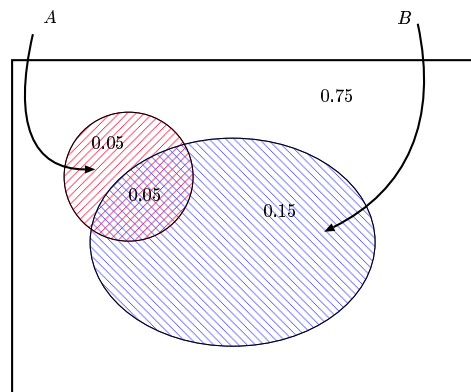
2.7 Enligt additionssatsen är

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

vilket ger

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.6 + 0.7 - 0.8 = 0.5.$$

2.8 Givet är $P(A) = 0.1$, $P(B) = 0.2$ och $P(A \cap B) = 0.05$. Ett Venn-diagram för händelserna kan se ut som:



Alltså,

a) $P(\{\text{Åtminstone ett av felen}\}) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.10 + 0.20 - 0.05 = 0.25.$

b) $P(\{A \text{ men ej } B\}) = P(A \cap B^*) = P(A) - P(A \cap B) = 0.10 - 0.05 = 0.05.$

c) $P(\{B \text{ men ej } A\}) = P(B \cap A^*) = P(B) - P(A \cap B) = 0.20 - 0.05 = 0.15.$

d) $P(\{\text{exakt ett av felen}\}) = P((B \cap A^*) \cup (A \cap B^*)) = P(B \cap A^*) + P(A \cap B^*) = 0.15 + 0.05 = 0.20.$

2.9 Om $A \cap B = \emptyset$ så är $P(A \cup B) = P(A) + P(B) = 0.6 + 0.7 = 1.3$ vilket inte är möjligt. Alltså, händelserna A och B är inte oförenliga (disjunkta).

2.10 Händelsen $A \cup B$ består av de disjunkta mängderna A och $A^* \cap B$, det vill säga $A \cup B = A \cup (A^* \cap B)$. Alltså är

$$P(A \cup B) = P(A) + P(A^* \cap B) = 0.5 + 0.1 = 0.6.$$

2.11 Händelsen A kan delas in i två disjunkta mängder: $A \cap B$ och $A \cap B^*$. Alltså är $P(A) = P(A \cap B) + P(A \cap B^*)$ vilket omformas till

$$P(A \cap B^*) = P(A) - P(A \cap B).$$

På samma sätt är $P(B \cap A^*) = P(B) - P(A \cap B)$, och eftersom $A \cap B^*$ och $B \cap A^*$ är disjunkta, så är

$$P((A \cap B^*) \cup (B \cap A^*)) = P(A \cap B^*) + P(B \cap A^*) = P(A) + P(B) - 2P(A \cap B).$$

2.12

a)

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= P(A_1 \cup (A_2 \cup \dots \cup A_n)) \\ &= P(A_1) + P(A_2 \cup \dots \cup A_n) - \underbrace{P(A_1 \cap (A_2 \cup \dots \cup A_n))}_{\geq 0} \\ &\leq P(A_1) + P(A_2 \cup \dots \cup A_n) \end{aligned}$$

på samma sätt

$$\leq P(A_1) + P(A_2) + P(A_3 \cup \dots \cup A_n)$$

och så vidare...

$$\leq P(A_1) + P(A_2) + \dots + P(A_n)$$

b)

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= \{\text{de Morgan}\} = P((A_1^* \cup \dots \cup A_n^*)^*) \\ &= 1 - P(A_1^* \cup \dots \cup A_n^*) \geq 1 - \sum_{i=1}^n P(A_i^*) = 1 - \sum_{i=1}^n (1 - P(A_i)). \end{aligned}$$

2.13 Vi väljer två bokstäver utan återläggning. Antalet sätt som dessa två bokstäver kan väljas på är $\binom{8}{2} = \frac{8 \cdot 7}{2 \cdot 1} = 28$ och vid val på måfå är alla dessa sätt lika sannolika. Enligt den klassiska sannolikhetsdefinitionen fås att sannolikheten kan bestämmas som

$$P(\text{Drar V och G}) = \frac{\text{Antal gynnsamma utfall}}{\text{Antal möjliga utfall}} = \frac{1}{28} \approx 0.0357.$$

2.14 Vi väljer tre kort utan återläggning. Antalet sätt som dessa tre kort kan väljas på är $\binom{52}{3} = \frac{52 \cdot 51 \cdot 50}{3 \cdot 2 \cdot 1} = 22100$ och vid val på måfå är alla dessa sätt lika sannolika.

a) Att välja tre kort så att alla tre korten är hjärter kan göras på $\binom{13}{3} = 286$ sätt. Enligt den klassiska sannolikhetsdefinitionen är

$$P(\text{Alla hjärter}) = \frac{\text{Antal gynnsamma utfall}}{\text{Antal möjliga utfall}} = \frac{\binom{13}{3}}{\binom{52}{3}} = \frac{11}{850} \approx 1.29\%.$$

b) Att välja tre kort så att inget kort är hjärter kan göras på $\binom{39}{3} = 9139$ sätt.

$$P(\text{Inga hjärter}) = \frac{\binom{39}{3}}{\binom{52}{3}} = \frac{703}{1700} \approx 41.35\%.$$

c) Att välja tre kort så att alla kort är ess kan göras på $\binom{4}{3} = 4$ sätt.

$$P(\text{Alla ess}) = \frac{\binom{4}{3}}{\binom{52}{3}} = \frac{1}{5525} \approx 0.0181\%.$$

2.15 Vi väljer två kulor utan återläggning. Antalet sätt som dessa två kulor kan väljas på är $\binom{7}{2} = \frac{7 \cdot 6}{2 \cdot 1} = 21$ och vid val på måfå är dessa 21 sätt lika sannolika.

Att välja två kulor så att en är svart och en är vit kan enligt multiplikationsprincipen göras på

$$\binom{\# \text{ sätt att välja en svart kula}}{\# \text{ sätt att välja en vit kula}} = \binom{4}{1} \binom{3}{1} = 4 \cdot 3 = 12 \text{ sätt.}$$

Enligt den klassiska sannolikhetsdefinitionen fås att

$$P(\text{Två kulor av olika färg}) = \frac{\binom{4}{1} \binom{3}{1}}{\binom{7}{2}} = \frac{12}{21} = \frac{4}{7} \approx 0.571.$$

b) Om urvalet görs med återläggning finns $7 \cdot 7 = 49$ sätt att välja två kulor på (med hänsyn till ordning). Enligt tidigare kan man på 12 sätt välja två kulor av olika färg oavsett ordning, det vill säga på $2 \cdot 4 \cdot 3 = 24$ sätt med hänsyn till ordning. Den sökta sannolikheten blir således

$$P(\text{Två kulor av olika färg}) = \frac{24}{49} \approx 0.490.$$

Alternativt kan sannolikheterna beräknas med hjälp av betingning på färgen på den först dragna kulan.



I fallet utan återläggning (till vänster) är sannolikheten för att de två kulorna har olika färg $\frac{4}{7} \cdot \frac{3}{6} + \frac{3}{7} \cdot \frac{4}{6} = \frac{24}{42} = \frac{4}{7}$ och i fallet med återläggning (till höger) är motsvarande sannolikhet $\frac{4}{7} \cdot \frac{3}{7} + \frac{3}{7} \cdot \frac{4}{7} = \frac{24}{49}$.

2.16 Mängden av alla stryktipsrader $\Omega = \{1, \times, 2\} \times \dots \times \{1, \times, 2\}$ har enligt multiplikationsprincipen $m = 3 \cdot 3 \cdot \dots \cdot 3 = 3^{13} = 1594323$ element vardera med sannolikhet $1/m$.

a) Låt A vara händelsen att man får 13 rätt. Antalet element i A är enligt multiplikationsprincipen

$$g = |A| = 1 \cdot 1 \cdot \dots \cdot 1 = 1^{13} = 1$$

$$\text{och } P(A) = g/m = 1/3^{13} = 3^{-13} \approx 6.272 \cdot 10^{-7}.$$

b) Låt B vara händelsen att de 12 första matcherna är rätt och den 13:e matchen fel. Antalet element i B ges av

$$g = |B| = 1 \cdot 1 \cdot \dots \cdot 1 \cdot 2 = 1^{12} \cdot 2 = 2$$

ty de första tolv matcherna kan endast tippas på ett sätt och sista matchen på 2 (felaktiga) sätt. Alltså $P(B) = g/m = 2/3^{13} \approx 1.255 \cdot 10^{-6}$.

c) Låt C vara händelsen att få "precis 12 rätt". Antalet stryktipsrader med tolv rätt där match i , $i = 1, \dots, 13$, är fel är

$$\underbrace{1 \cdot \dots \cdot 1}_{i-1 \text{ st}} \cdot 2 \cdot \underbrace{1 \cdot \dots \cdot 1}_{13-i \text{ st}} = 1^{12} \cdot 2 = 2,$$

eftersom man kan tippa match i fel på två sätt. Med 13 möjliga värden på i så är antalet stryktipsrader med exakt 12 rätt $|C| = 13 \cdot 2 = 26$ och $P(C) = g/m = 26/3^{13} = 26 \cdot 3^{-13} \approx 1.631 \cdot 10^{-5}$.

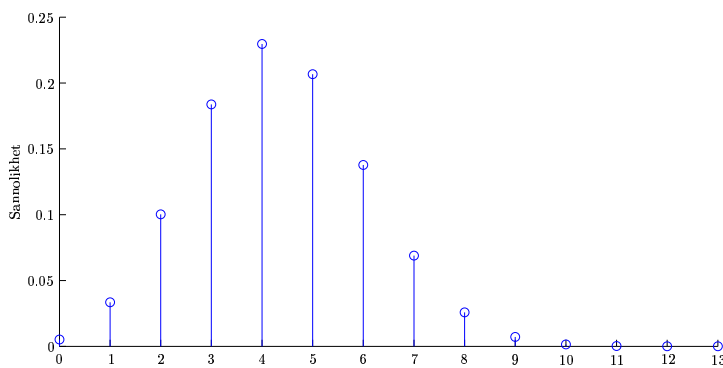
Generellt, låt D_k vara händelsen "precis k rätt". Enligt multiplikationsprincipen är antalet utfall i D_k ,

$$|D_k| = \binom{\# \text{ sätt att välja ut } k \text{ matcher}}{\# \text{ sätt att tippa } k \text{ matcher rätt}} \binom{\# \text{ sätt att tippa } 13 - k \text{ matcher fel}}{\# \text{ sätt att tippa } 13 - k \text{ matcher rätt}} = \binom{13}{k} \cdot 1^k \cdot 2^{13-k}$$

så sannolikheten för att en stryktipsrad har exakt k rätt ges av

$$P(\text{precis } k \text{ rätt}) = \frac{\binom{13}{k} 1^k 2^{13-k}}{3^{13}} = \binom{13}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{13-k}$$

för $k = 0, 1, \dots, 13$.



Sannolikheterna för att få k rätt på stryktipset med en på måfå ifylld rad.

2.17 Sannolikheten att det bland n personer inte finns någon gemensam födelsedag är

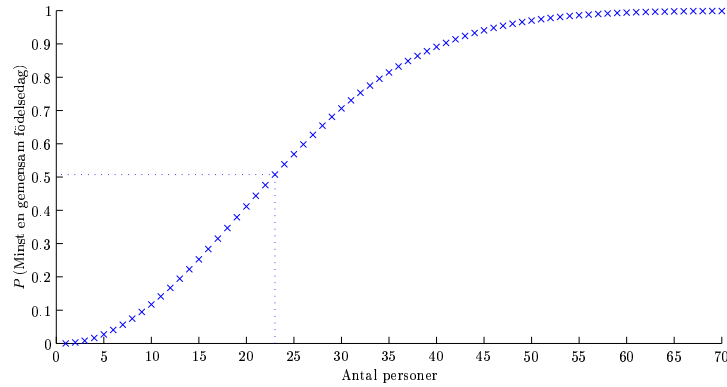
$$P(\text{Ingen gemensam}) = \frac{\text{Antalet sätt vi kan välja ut } n \text{ födelsedagar utan par}}{\text{Antalet sätt vi kan välja ut } n \text{ födelsedagar}}$$

Under ett antagande om att varje år har 365 dagar och att födelsedagar är likformigt fördelade över åren kan med multiplikationsprincipen bestämma sannolikheten till

$$P(\text{Ingen gemensam}) = \frac{365 \cdot 364 \cdot \dots \cdot (365 - n + 1)}{365 \cdot 365 \cdot \dots \cdot 365} = \frac{365!}{365^n (365 - n)!}.$$

Sannolikheten för minst att det bland n personer finns minst en gemensam födelsedag är således

$$P(\text{Minst en gemensam}) = 1 - P(\text{Ingen gemensam}) = 1 - \frac{365!}{365^n (365 - n)!}.$$



För olika värden på n bestäms sannolikheten till

n	2	5	10	20	23	30	40
$P(\text{Minst en gemensam})$.0027	.0271	.1169	.4114	.5073	.7063	.8912,

och man ser att för 23 eller fler personer är sannolikheten större än 50%.

2.18 Slumpförsöket består av att dra 5 kort utan återläggning på måfå ur en kortlek om 52 kort. Utfallsrummet består då av de $m = \binom{52}{5} = 2598960$ sätt som detta kan göras på. Antalet utfall som motsvarar en hand med korten...

a) ess, kung, dam, knekt, tio i samma färg, är enligt multiplikationsprincipen

$$g = (\# \text{ färger}) \cdot \binom{\# \text{ följder}}{\text{ess}, \dots, \text{tio}} = 4 \cdot 1$$

så sannolikheten är $g/m = 4/\binom{52}{5}$.

b) fem kort i följd i samma färg, är enligt multiplikationsprincipen

$$g = (\# \text{ färger}) \cdot \binom{\# \text{ följder om}}{\text{fem kort}} = 4 \cdot 10 = 40$$

så sannolikheten är $g/m = 40/\binom{52}{5}$.

c) fem kort i samma färg, är enligt multiplikationsprincipen

$$g = (\# \text{ färger}) \cdot \binom{\# \text{ sätt att}}{\text{välja fem kort}} = 4 \cdot \binom{13}{5} = 5148$$

så sannolikheten är $g/m = 4 \cdot \binom{13}{5} / \binom{52}{5}$.

2.19 Slumpförsöket består av att dra 13 kort utan återläggning på måfå ur en kortlek om 52 kort. Utfallsrummet består då av de $m = \binom{52}{13} = 63501359600$ sätt som detta kan göras på. Antalet utfall som motsvarar en hand med korten...

a) $5\spadesuit, 3\heartsuit, 3\diamondsuit, 2\clubsuit$ är enligt multiplikationsprincipen

$$\begin{aligned} g &= \binom{\# \text{ sätt att}}{\text{välja } 5 \spadesuit} \cdot \binom{\# \text{ sätt att}}{\text{välja } 3 \heartsuit} \cdot \binom{\# \text{ sätt att}}{\text{välja } 3 \diamondsuit} \cdot \binom{\# \text{ sätt att}}{\text{välja } 2 \clubsuit} \\ &= \binom{13}{5} \binom{13}{3} \binom{13}{3} \binom{13}{2} = 1287 \cdot 286 \cdot 286 \cdot 78 = 8211173256 \end{aligned}$$

så sannolikheten är $g/m = 0.01293$.

b) fördelningen 5,3,3,2 på godtyckliga (distinkta) färger. Enligt ovan är

$$\binom{\# \text{ sätt att välja 5 av färg 1}}{\binom{\# \text{ sätt att välja 3 av färg 2}}{\binom{\# \text{ sätt att välja 3 av färg 3}}{\binom{\# \text{ sätt att välja 2 av färg 4}}}} = 8211173256$$

så det som återstår är att bestämma på hur många olika sätt man kan fördela ♡, ♠, ♣ och ♣ över färg 1–4. Välj först vilka färger vi skall plocka tre kort av. Det kan göras på $\binom{4}{2} = 6$ sätt. För vart och ett av de sätten skall vi bestämma den färg som vi skall plocka 5 kort av. Vi har två färger kvar så detta kan göras på $\binom{2}{1} = 2$ sätt. Den sista färgen kan bara väljas på ett sätt. Det totala antalet sätt som färgerna kan fördelas är

$$\binom{4}{2} \cdot 2 \cdot 1 = 12$$

och sannolikheten är $12 \cdot (\text{svaret i a}) = 0.15516$.

2.20

Alternativ 1: Betrakta den hög där ruter kung ligger. Sannolikheten att hjärter kung ligger i samma hög är $1/3$ (ett kort av tre möjliga), dvs. med sannolikhet $2/3$ ligger de i olika högar.

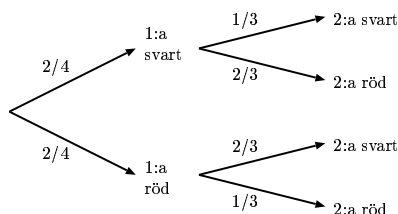
Alternativ 2: (Hypergeometrisk fördelning.) Av $N = 4$ kort är $s = 2$ svarta och $v = 2$ röda. För att välja ut korten i den ena högen väljer man ut $n = 2$ kort av N stycken, vilket kan göras på $\binom{N}{n} = \binom{4}{2} = \frac{4 \cdot 3}{2 \cdot 1} = 6$ sätt. Av dessa är det

$$\binom{s}{1} \binom{v}{1} = \binom{2}{1} \binom{2}{1} = 2 \cdot 2 = 4$$

sätt som ger exakt ett svart kort, vilket medför att de två röda korten ligger i varsin hög. Den sökta sannolikheten är således

$$\frac{\binom{2}{1} \binom{2}{1}}{\binom{4}{2}} = \frac{2}{3}.$$

Alternativ 3: (Formell betingning) De två korten som skall läggas i en av högarna väljes enligt nedanstående träd diagram:



Sannolikheten för att få två kort av olika färg är således

$$\frac{2}{4} \cdot \frac{2}{3} + \frac{2}{4} \cdot \frac{2}{3} = \frac{2}{3}.$$

2.21 Av $N = 100$ distinkta enheter är $s = 6$ defekta. Om man väljer ut $n = 5$ på måfå utan återläggning så är enligt multiplikationsprincipen

$$\begin{aligned} P(2 \text{ defekta}) &= \frac{\binom{\# \text{ sätt att välja 2 bland de defekta}}{\binom{\# \text{ sätt att välja 5 hela}}{\binom{5-2 \text{ bland de}}{\binom{hela}}}}} \\ &= \frac{\binom{6}{2} \binom{94}{3}}{\binom{100}{5}} = \frac{15 \cdot 134044}{75287520} = \frac{33511}{1254792} \approx 0.026706. \end{aligned}$$

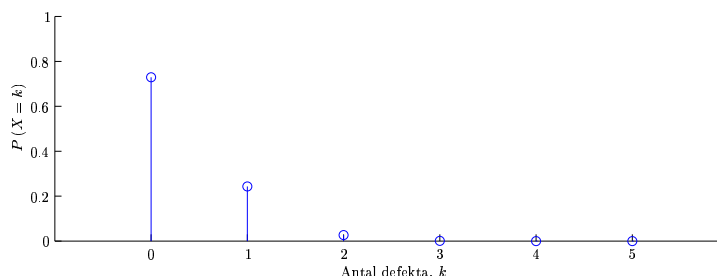
På samma sätt kan man räkna ut

$$P(0 \text{ eller } 1 \text{ defekt}) = \frac{\binom{6}{0} \binom{94}{5} + \binom{6}{1} \binom{94}{4}}{\binom{100}{5}} = \frac{54891018 + 18297006}{75287520} = \frac{435643}{448140} \approx 0.97211.$$

Om man låter $\{X = k\}$ beteckna händelsen att det finns k defekta bland de utvalda så kan man på samma sätt bestämma sannolikheten för händelserna

$$P(X = k) = \frac{\binom{\text{\#sätt att välja } k \text{ bland de defekta}}{\binom{\text{\#sätt att välja } 5 - k \text{ bland de hela}}{\binom{6}{k} \binom{94}{5 - k}}}{\binom{100}{5}}.$$

för $k = 0, 1, \dots, 5$



2.22 Efter ha tagit tre hjärter ur kortleken finns $52 - 3 = 49$ kort kvar varav $13 - 3 = 10$ är hjärter. Den betingade sannolikheten att det fjärde kortet inte är ett hjärter är

$$P(\text{Fjärde kortet är inte hjärter} | \text{tre första korten är hjärter}) = \frac{39}{49} \approx 0.796.$$

b) Låt A vara händelsen att det fjärde kortet är en spader och B händelsen att de tre första korten är hjärter.

Notera att det inte är en betingad sannolikhet som söks. Av de kvarvarande 49 korten är 13 stycken spader så sannolikheten är

$$P(\text{Fjärde kortet är spader} | \text{tre första korten är hjärter}) = P(A|B) = \frac{13}{49} \approx 0.265$$

det vill säga en tredjedel av svaret i a.

Den sökta sannolikheten är

$$P(A \cap B) = P(A|B) P(B)$$

där

$$P(B) = P(\text{tre första korten är hjärter}) = \frac{13 \cdot 12 \cdot 11}{52 \cdot 51 \cdot 50} = \frac{11}{850}$$

så

$$P(\{\text{Fjärde kortet är spader}\} \cap \{\text{tre första korten är hjärter}\}) = P(A|B) P(B) = \frac{13}{49} \cdot \frac{11}{850} = \frac{143}{41650}.$$

2.23 Låt H_1 , H_2 och H_3 beteckna händelserna att instrument 1, 2 eller 3 väljes. Då är $P(H_k) = \frac{1}{3}$, $k = 1, 2, 3$, och de tre händelserna utgör en partition av utfallsrummet. Låt A vara händelsen att valt instrument fungerar. Enligt lagen om total sannolikhet är

$$P(A) = P(A|H_1) P(H_1) + P(A|H_2) P(H_2) + P(A|H_3) P(H_3) = 0.9 \cdot \frac{1}{3} + 0.8 \cdot \frac{1}{3} + 0.4 \cdot \frac{1}{3} = 0.7.$$

b) Med Bayes formel erhålles

$$P(H_1|A) = \frac{P(H_1 \cap A)}{P(A)} = \frac{P(A|H_1)P(H_1)}{P(A)} = \frac{0.9 \cdot \frac{1}{3}}{0.7} = \frac{3}{7}.$$

På samma sätt fås

$$P(H_2|A) = \frac{P(A|H_2)P(H_2)}{P(A)} = \frac{0.8 \cdot \frac{1}{3}}{0.7} = \frac{8}{21}$$

och

$$P(H_3|A) = \frac{P(A|H_3)P(H_3)}{P(A)} = \frac{0.4 \cdot \frac{1}{3}}{0.7} = \frac{4}{21}.$$

Notera att dessa tre sannolikheter summeras till 1.

2.24 Av de 11 frukterna är 3 giftiga och 8 ogiftiga. Låt B vara händelsen att hunden får en ogiftig frukt. Hundens frukt väljes på måfå och således är $P(B) = 8/11$.

Givet att hunden klarar sig finns 10 frukter kvar varav 3 är giftiga och 7 ogiftiga. Sannolikheten att Per får alla giftiga frukter är

$$\frac{\binom{\text{\#sätt att välja 3}}{\text{bland de giftiga}} \binom{\text{\#sätt att välja}}{4-3 \text{ bland de}}{\text{ogiftiga}}}{\text{\#sätt att välja 4 bland alla}} = \frac{\binom{3}{3} \binom{7}{1}}{\binom{10}{4}} = \frac{1 \cdot 7}{\frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1}} = \frac{1}{30}.$$

Sannolikheten att Pål får alla giftiga frukter är

$$\frac{\binom{\text{\#sätt att välja 3}}{\text{bland de giftiga}} \binom{\text{\#sätt att välja}}{6-3 \text{ bland de}}{\text{ogiftiga}}}{\text{\#sätt att välja 6 bland alla}} = \frac{\binom{3}{3} \binom{7}{3}}{\binom{10}{6}} = \frac{1 \cdot \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1}}{\frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1}} = \frac{1}{6}.$$

Med A som händelsen att både Per och Pål får minst en giftig frukt var är

$$\begin{aligned} P(A|B) &= 1 - P(\{\text{Per får alla}\} \cup \{\text{Pål får alla}\}|B) = 1 - P(\text{Per får alla}|B) - P(\text{Pål får alla}|B) \\ &= 1 - \frac{1}{30} - \frac{1}{6} = \frac{4}{5}. \end{aligned}$$

c) Slutligen

$$P(A \cap B) = P(A|B)P(B) = \frac{4}{5} \cdot \frac{8}{11} = \frac{32}{55}.$$

2.25 Låt H vara händelsen att det var två M som föll ned. Då är $P(H) = 1/\binom{5}{2} = 1/10$. Om två M faller ned står det fortfarande MALMÖ efter uppsättning. Om det inte är två M som faller ned står det MALMÖ på skylten efter uppsättning med sannolikhet $1/2$. Enligt lagen om total sannolikhet är

$$P(\text{MALMÖ}) = P(\text{MALMÖ}|H)P(H) + P(\text{MALMÖ}|H^*)P(H^*) = 1 \cdot \frac{1}{10} + \frac{1}{2} \cdot \frac{9}{10} = \frac{11}{20}.$$

2.26 Händelsen att familjen har minst ett barn av vardera kön givet att det äldsta barnet är en pojke är händelsen att de tre andra barnen är en pojke och två flickor. Varje sådan tilldelning av kön har sannolikhet $p(1-p)^2$ enligt oberoendet. Att välja ut vilket barn som blir pojke kan göras på $\binom{3}{1} = 3$ sätt. Den sökta sannolikheten är således $3p(1-p)^2$.

b) Låt B vara händelsen att familjen har minst en pojke. Då är B^* händelsen att familjen har enbart flickor och enligt oberoendet $P(B^*) = (1-p)^4$, det vill säga $P(B) = 1 - (1-p)^4$.

Låt A vara händelsen att familjen har två pojkar och två flickor. Varje sekvens av två pojkar och två flickor har sannolikhet $p^2(1-p)^2$ på grund av oberoendet och det finns $\binom{4}{2} = 6$ sådana sekvenser (könstilldelningar). Således är $P(A) = 6p^2(1-p)^2$.

Nu är $A \subset B$ och

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{6p^2(1-p)^2}{1 - (1-p)^4}.$$

2.27

a) Med definitionen av betingad sannolikhet har man att

$$P(\text{andra sidan röd} | \text{sedd sida röd}) = \frac{P(\text{andra sidan röd} \cap \text{sedd sida röd})}{P(\text{sedd sida röd})} = \frac{1/3}{1/2} = \frac{2}{3},$$

eftersom täljaren är sannolikheten för händelsen att man ser det helröda kortet och nämnaren fås av att hälften av kortsidorna är röda.

b) Det är inte samma chans för båda fallen. Av ovanstående följer att

$$P(\text{rödröd} | \text{sedd sida röd}) = P(\text{andra sidan röd} | \text{sedd sida röd}) = 2/3$$

medan

$$P(\text{rödvit} | \text{sedd sida röd}) = P(\text{andra sidan vit} | \text{sedd sida röd}) = 1/3.$$

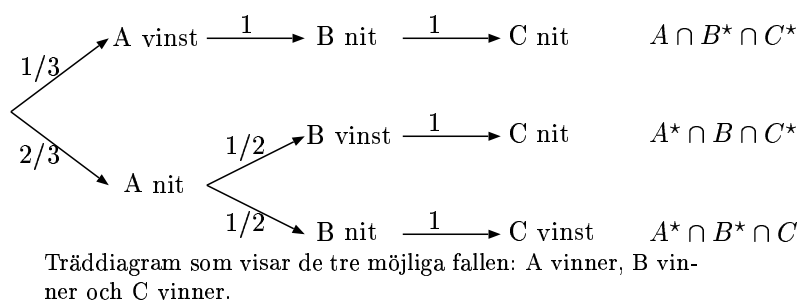
2.28 Låt A , B och C vara händelsen att person A, B och C får vinstlotten. Notera att händelserna är disjunkta och $P(A \cup B \cup C) = 1$. Eftersom A tar första lotten är $P(A) = 1/3 = P(A \cap B^* \cap C^*)$.

$$P(B) = P(B|A)P(A) + P(B|A^*)P(A^*) = 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

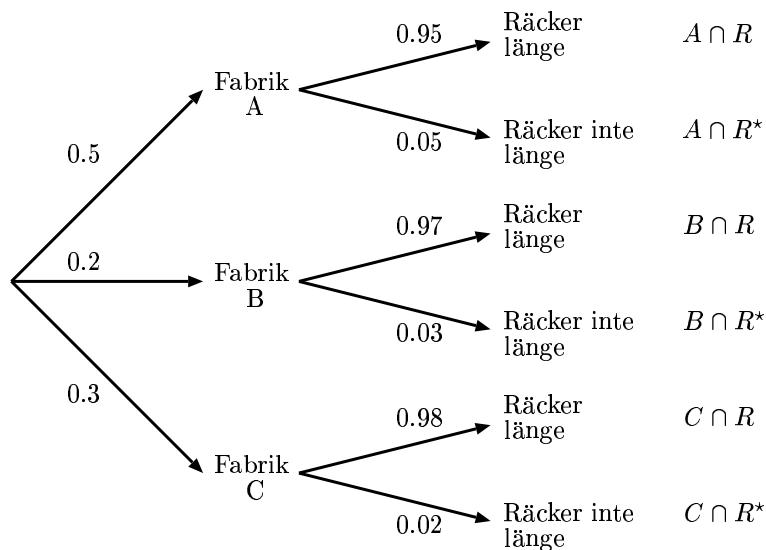
eftersom givet A^* finns två lotter kvar och $P(B|A^*)$ är således en halv. Slutligen,

$$P(C) = P(C|B)P(B) + P(C|B^*)P(B^*) = 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Resonemanget illustreras enkelt med ett träd-diagram:



2.29 Låt A , B och C beteckna händelserna att ett på måfå valt batteri kommer från fabrik A, B eller C. Vidare, låt R beteckna händelsen att valt batteri har en lång livslängd. Enklart illustreras sambandet mellan händelserna och deras betingade sannolikheter i ett träd-diagram såsom det nedan. Sannolikheten för tex. händelsen att man väljer ett batteri från fabrik A som räcker länge, $P(A \cap R)$, kan beräknas som $P(R|A)P(A)$, dvs. genom att multiplicera grenarnas sannolikheter.



Allts ,

$$P(A \cap R) = P(R|A)P(A) = 0.95 \cdot 0.5 = 0.475$$

$$P(B \cap R) = P(R|B)P(B) = 0.97 \cdot 0.2 = 0.194$$

$$P(C \cap R) = P(R|C)P(C) = 0.98 \cdot 0.3 = 0.294$$

och

$$P(R) = P((A \cap R) \cup (B \cap R) \cup (C \cap R)) = P(A \cap R) + P(B \cap R) + P(C \cap R) = 0.963.$$

Vidare s   r

$$P(A|R) = \frac{P(A \cap R)}{P(R)} = \frac{0.475}{0.963} = 0.49325$$

och

$$P(A|R^*) = \frac{P(A \cap R^*)}{P(R^*)} = \frac{P(R^*|A)P(A)}{1 - P(R)} = \frac{0.05 \cdot 0.5}{0.037} = 0.6757.$$

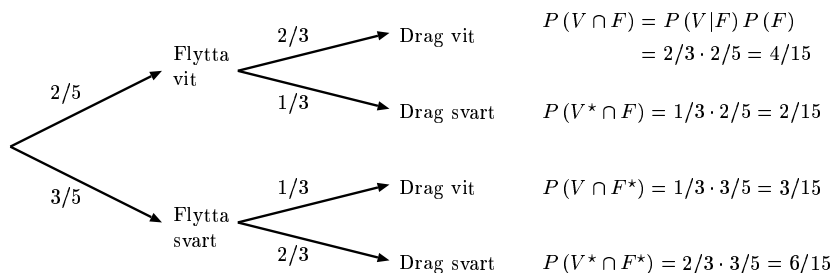
2.30 L t F vara h ndelsen "den flyttade kulan  r vit" och V h ndelsen "den dragna kulan  r vit".

a) D   r F^* h ndelsen "flyttad kula svart" och

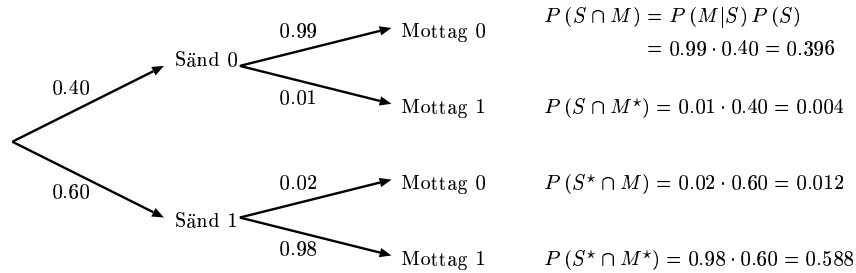
$$\begin{aligned} P(V) &= P((V \cap F) \cup (V \cap F^*)) = P(V \cap F) + P(V \cap F^*) \\ &= P(V|F)P(F) + P(V|F^*)P(F^*) = \frac{2}{3} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{3}{5} = \frac{7}{15}. \end{aligned}$$

b) Vi s ker nu $P(F^*|V)$.

$$P(F^*|V) = \frac{P(F^* \cap V)}{P(V)} = \frac{P(V|F^*)P(F^*)}{P(V)} = \frac{\frac{1}{3} \cdot \frac{3}{5}}{\frac{7}{15}} = \frac{3}{7}.$$



2.31 Låt S vara händelsen att en skickad bit är en 0:a och M vara händelsen att en 0:a mottagits. Beroendet mellan S och M ges av trädidiagrammet nedan.



De sökta sannolikheterna är (kolla figuren!)

$$P(S^*|M^*) = \frac{P(S^* \cap M^*)}{P(M^*)} = \frac{P(M^*|S^*)P(S^*)}{P(M^*|S)P(S) + P(M^*|S^*)P(S^*)} = \frac{0.588}{0.004 + 0.588} = 0.9932$$

och

$$P(\text{fel}) = P((S \cap M^*) \cup (S^* \cap M)) = 0.004 + 0.012 = 0.016.$$

2.32 Med de Morgan

$$\begin{aligned} P(A^* \cap B^*) &= P((A \cup B)^*) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) \\ &= (A \text{ och } B \text{ oberoende}) = 1 - P(A) - P(B) + P(A)P(B) = (1 - P(A))(1 - P(B)) \\ &= P(A^*)P(B^*) \end{aligned}$$

det vill säga A^* och B^* är oberoende. Således

$$P(A^* \cap B^*) = P(A^*)P(B^*) = (1 - 0.1)(1 - 0.05) = 0.855.$$

2.33 Låt A , B och C vara händelserna att familj A, B respektive C kommer där $P(A) = 0.8$, $P(B) = 0.6$ och $P(C) = 0.9$.

a) Händelsen att alla kommer är $A \cap B \cap C$ och med oberoendet fås

$$P(\text{Alla kommer}) = P(A \cap B \cap C) = P(A)P(B)P(C) = 0.8 \cdot 0.6 \cdot 0.9 = 0.432.$$

b) På samma sätt

$$P(\text{Ingen kommer}) = P(A^* \cap B^* \cap C^*) = P(A^*)P(B^*)P(C^*) = 0.2 \cdot 0.4 \cdot 0.1 = 0.008.$$

c) Komplementet till händelsen {minst en kommer} är att ingen kommer. Således är

$$P(\text{Minst en kommer}) = 1 - P(\text{Ingen kommer}) = 0.992.$$

2.34 Om tärningen har 6 sidor och har samma sannolikhet för varje sida är $P(A) = 2/6 = 1/3$. Genom att betinga på utfallet i första tärningskastet kan man utnyttja lagen om total sannolikhet för att bestämma $P(B)$:

$$\begin{aligned} P(B) &= P(B|1:a)P(1:a) + P(B|2:a)P(2:a) + P(B|3:a)P(3:a) + P(B|4:a)P(4:a) \\ &\quad + P(B|5:a)P(5:a) + P(B|6:a)P(6:a) = \frac{1}{6} \cdot \frac{1}{6} + \frac{2}{6} \cdot \frac{1}{6} + \frac{3}{6} \cdot \frac{1}{6} + \frac{4}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} = \frac{21}{36}. \end{aligned}$$

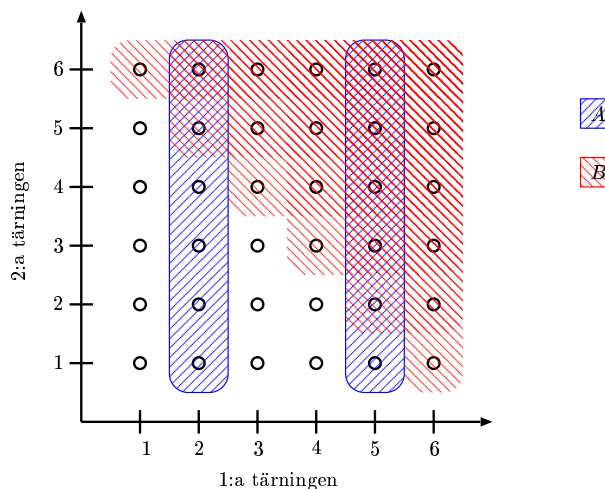
Händelsen $A \cap B$ är händelsen {slå en 2:a följt av en 5:a eller 6:a} eller {slå en 5:a följt av minst en 2:a}, det vill säga

$$P(A \cap B) = \frac{1}{6} \cdot \frac{2}{6} + \frac{1}{6} \cdot \frac{5}{6} = \frac{7}{36}.$$

Eftersom

$$P(A \cap B) = \frac{7}{36} = \frac{1}{3} \cdot \frac{21}{36} = P(A) P(B)$$

så är händelserna oberoende. Det ovanstående fås även direkt ur figuren nedan:



2.35

a) Med data:

	k	t	
s	1	1	2
v	1	1	2
	2	2	4

fås $P(S) = 2/4 = 1/2$, $P(T) = 2/4 = 1/2$ och $P(S \cap T) = 1/4$. Alltså är

$$P(S \cap T) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(S) P(T)$$

och händelserna är oberoende.

b) Med data:

	k	t	
s	1	10	11
v	10	1	11
	11	11	22

fås $P(S) = 11/22 = 1/2$, $P(T) = 11/22 = 1/2$ och $P(S \cap T) = 10/22$. Alltså är

$$P(S \cap T) = \frac{10}{22} \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(S) P(T)$$

och händelserna är inte oberoende. Givet information att det dragna föremålet är säg svart ökar sannolikheten att det också är en tärning. $P(T|S) = 10/11$.

2.36 Sannolikheten att en tillverkad komponent har minst ett av felen

Alternativ 1: Händelsen "något av felen" = $A \cup B \cup C$ och

$$\begin{aligned} P(A \cup B \cup C) &= \{\text{de Morgan}\} = 1 - P(A^* \cap B^* \cap C^*) = \{\text{oberoende}\} \\ &= 1 - P(A^*) P(B^*) P(C^*) = 1 - (1 - 0.20)(1 - 0.05)(1 - 0.10) \\ &= 1 - 0.684 = 0.316. \end{aligned}$$

Alternativ 2: Händelsen "något av felen" = $A \cup B \cup C$ och

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) = \{\text{oberoende}\} \\ &= P(A) + P(B) + P(C) - P(A)P(B) - P(A)P(C) - P(B)P(C) \\ &\quad + P(A)P(B)P(C) \\ &= 0.2 + 0.3 + 0.5 - 0.2 \cdot 0.3 - 0.2 \cdot 0.5 - 0.3 \cdot 0.5 + 0.2 \cdot 0.3 \cdot 0.5 = 0.316. \end{aligned}$$

Alternativ 3: Låt X beskriva antalet fel hos produkten. $\{X = 0\} = A^* \cap B^* \cap C^*$ så utnyttjandes oberoendet

$$P(X = 0) = P(A^*) P(B^*) P(C^*) = 0.684$$

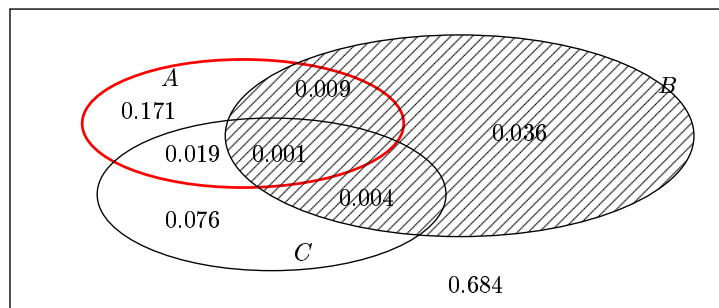
och på samma sätt fås

$$P(X = 1) = P(A^*) P(B^*) P(C) + P(A^*) P(B) P(C^*) + P(A) P(B^*) P(C^*) = 0.283$$

$$P(X = 2) = P(A) P(B) P(C^*) + P(A) P(B^*) P(C) + P(A^*) P(B) P(C) = 0.032$$

$$P(X = 3) = P(A) P(B) P(C) = 0.001$$

Vi söker $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.684 = 0.316$.



Venn-diagram där händelserna motsvarande felen A , B och C är oberoende.

2.37 Låt A vara händelsen att det dragna kortet är hjärter och B händelsen att det dragna kortet är ess. Vi får omedelbart att $P(A) = 13/52 = 1/4$ och $P(B) = 4/52 = 1/13$. Händelsen $A \cap B$ är händelsen att det dragna kortet är hjärter ess och vi får då att $P(A \cap B) = 1/52$. Det innebär att $P(A \cap B) = P(A)P(B)$ varför A och B är oberoende.

b) Definiera A och B som i a) ovan. Vi får då att $P(A) = 13/48$ och $P(B) = 4/48 = 1/12$. Vidare ser vi att $P(A \cap B) = P(\text{dragna kortet är hjärter ess}) = 1/48$ varför vi *inte* har att $P(A \cap B) = P(A)P(B)$. Händelserna A och B är ej oberoende.

2.38 Låt A och B vara två händelser med $P(A) > 0$ och $P(B) > 0$.

a) Om A och B är oförenliga (disjunkta), $A \cap B = \emptyset$, så är

$$0 = P(\emptyset) = P(A \cap B) \neq \underbrace{P(A)}_{>0} \underbrace{P(B)}_{>0} > 0$$

och händelserna är *inte* oberoende.

b) Om A och B är oberoende så är

$$P(A \cap B) = \underbrace{P(A)}_{>0} \underbrace{P(B)}_{>0} > 0$$

så $A \cap B \neq \emptyset$ och A och B är ej oförenliga.

2.39 Låt K_1, \dots, K_n vara de oberoende händelserna att komponenter $1, \dots, n$ fungerar en given tid. $P(K_i) = p_i$.

a) Då gäller för ett *seriesystem*

$$P(\text{Syst. fungerar}) = P(K_1 \cap \dots \cap K_n) = \{\text{ober.}\} = P(K_1) \cdots P(K_n) = p_1 \cdots p_n.$$

b) För ett *parallellsystem* har vi att

$$\begin{aligned} P(\text{Syst. fungerar}) &= P(K_1 \cup \dots \cup K_n) = 1 - P(K_1^* \cap \dots \cap K_n^*) = \{\text{ober.}\} \\ &= 1 - P(K_1^*) \cdots P(K_n^*) = 1 - (1 - p_1) \cdots (1 - p_n). \end{aligned}$$

c) Med $n = 4$ och $p_i = 0.90$ får man

$$\begin{aligned} P(\text{Seriesystem fungerar}) &= 0.90^4 = 0.6561 \\ P(\text{Parallellsystem fungerar}) &= 1 - (1 - 0.90)^4 = 1 - 10^{-4}. \end{aligned}$$

2.40 (Se även uppgift 2.39.) För två komponenter, låt A och B vara de oberoende händelserna att respektive komponent fungerar. Notera att med två komponenter är händelserna $\{\text{seriesystem fungerar}\} = A \cap B$ och $\{\text{parallellsystem fungerar}\} = A \cup B = (A^* \cap B^*)^*$.

a) Händelsen $\{\text{seriesystem fungerar}\} = A \cap B$ så med oberoendet

$$P(\{\text{seriesystem fungerar}\}) = P(A \cap B) = P(A)P(B) = 0.9 \cdot 0.8 = 0.72.$$

b) Komponentredundans motsvaras av en seriekoppling av två parallellsystem där parallellsystem 1 fungerar med sannolikhet

$$\begin{aligned} P(\{\text{parallellsystem 1 fungerar}\}) &= P(A_1 \cup A_2) = 1 - P(A_1^* \cap A_2^*) = 1 - P(A_1^*)P(A_2^*) \\ &= 1 - (1 - P(A_1))(1 - P(A_2)) = 1 - (1 - 0.9)(1 - 0.9) = 0.99. \end{aligned}$$

Motsvarande för parallellsystem 2 är

$$\begin{aligned} P(\{\text{parallellsystem 2 fungerar}\}) &= P(B_1 \cup B_2) = 1 - P(B_1^* \cap B_2^*) = 1 - P(B_1^*)P(B_2^*) \\ &= 1 - (1 - P(B_1))(1 - P(B_2)) = 1 - (1 - 0.8)(1 - 0.8) = 0.96. \end{aligned}$$

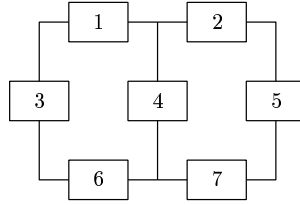
Seriekopplingen av dessa två system har funktionssannolikhet

$$\begin{aligned} P(\{\text{system fungerar}\}) &= P(\{\text{parallells. 1 fungerar}\} \cap \{\text{parallells. 2 fungerar}\}) \\ &= 0.99 \cdot 0.96 = 0.9504. \end{aligned}$$

c) Systemredundans innebär en parallellkoppling av två seriesystemen, där varje seriesystem har funktionssannolikhet 0.72. Parallellkopplingen har funktionssannolikhet

$$\begin{aligned} P(\{\text{system fungerar}\}) &= P(\{\text{seriesystem 1 fungerar}\} \cup \{\text{seriesystem 2 fungerar}\}) \\ &= 1 - (1 - P(\{\text{series. 1 fung}\}))(1 - P(\{\text{series 2 fung}\})) \\ &= 1 - (1 - 0.72)^2 = 0.9216. \end{aligned}$$

2.41 Numrera reläerna enligt:



Låt K_i , $i = 1, \dots, 7$, beteckna händelsen att relä i är tillslaget. Händelserna K_i , $i = 1, \dots, 7$ är oberoende och låt $p_i = P(K_i)$.

Låt

$$\begin{aligned} A &= K_3, \\ B &= K_1 \cap K_4 \cap K_6 \\ C &= K_1 \cap K_2 \cap K_5 \cap K_7 \cap K_6. \end{aligned}$$

Om någon av händelserna A , B eller C inträffar är punkterna förbundna med varandra. Alltså, punkterna är förbundna med sannolikhet

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C) \\ &= p_3 + p_1 p_4 p_6 + p_1 p_2 p_5 p_7 p_6 - p_3 p_1 p_4 p_6 - p_3 p_1 p_2 p_5 p_7 p_6 - p_1 p_4 p_6 p_2 p_5 p_7 + \\ &\quad p_1 p_2 p_3 p_4 p_5 p_6 p_7 = p + p^3 - p^4 + p^5 - 2p^6 + p^7. \end{aligned}$$

2.42 Spelet avslutas när tavlan träffats två gånger. Händelsen att det är samma person som skjutit båda skotten är händelsen att efter första träffen sker det ett udda antal missar följt av en träff.

Alltså,

$$\begin{aligned} P(\text{Samma person}) &= \sum_{k=0}^{\infty} (1-p)^{2k+1} p = (1-p)p \sum_{k=0}^{\infty} [(1-p)^2]^k = (1-p)p \frac{1}{1-(1-p)^2} \\ &= \frac{1-p}{2-p}. \end{aligned}$$

Notera att detta är sannolikheten att en $\text{ffg}(p)$ -fördelad stokastisk variabel antar ett jämnt värde.

2.43

Alternativ 1: En kortlek kan delas i 4 lika stora högar, det vill säga med 13 kort i varje hög, på

$$\binom{52}{13} \binom{39}{13} \binom{26}{13} \binom{13}{13}$$

sätt. Att välja ut den första högen så att den innehåller ett ess kan göras på $\binom{48}{12} \binom{4}{1}$ sätt. Efter detta kan den andra högen väljas på $\binom{36}{12} \binom{3}{1}$ sätt så att den innehåller ett ess. Den tredje högen väljes analogt på $\binom{24}{12} \binom{2}{1}$ sätt och den fjärde på $\binom{12}{12} \binom{1}{1}$ sätt.

Sålunda blir sannolikheten

$$P(\text{Ess i varje hög}) = \frac{\binom{48}{12} \binom{4}{1} \binom{36}{12} \binom{3}{1} \binom{24}{12} \binom{2}{1} \binom{12}{12} \binom{1}{1}}{\binom{52}{13} \binom{39}{13} \binom{26}{13} \binom{13}{13}} = \frac{13^4}{\binom{52}{4}} = \frac{2197}{20825} \approx 0.1055$$

Alternativ 2: Vi tar hänsyn till kortens ordning i högarna. Vi tänker oss 52 positioner (lådor) där låda 1–13 motsvarar hög 1, lådor 14–26 hög 2, 27–39 hög 3 och slutligen 40–52 den sista högen. Det första esset kan läggas ut på 52 sätt. När det lagts ut är de 12 andra lådorna knutna till samma hög blockerade så ess två kan läggas ut på 39 sätt. För var och ett av dessa kan sedan ess tre läggas ut på 26 sätt och slutligen det sista esset på 13 sätt. Sannolikheten är därför

$$\frac{52 \cdot 39 \cdot 26 \cdot 13}{52 \cdot 51 \cdot 50 \cdot 49} = \frac{(4 \cdot 13) \cdot (3 \cdot 13) \cdot (2 \cdot 13) \cdot (1 \cdot 13)}{52! / 48!} = \frac{13^4}{\frac{52!}{4!48!}} = \frac{13^4}{\binom{52}{4}}.$$

2.44 Låt B vara händelsen att bridgehanden innehåller minst ett ess och A händelsen att bridgehanden innehåller exakt ett ess.

En bridgehand om 13 kort kan väljas på $\binom{52}{13}$ sätt. Om man inte har några ess är de 13 korten valda bland de $52 - 4 = 48$ kort som inte är ess. En bridgehand utan ess kan således väljas på $\binom{48}{13}$ sätt och om bridgehanden väljes på måfå är

$$P(B) = P(\text{minst ett ess}) = 1 - P(\text{Inget ess}) = 1 - \frac{\binom{48}{13}}{\binom{52}{13}}.$$

Eftersom $A \subset B$ och

$$P(A) = P(\text{exakt ett ess}) = \frac{\binom{4}{1} \binom{48}{12}}{\binom{52}{13}}$$

så är

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\binom{4}{1} \binom{48}{12}}{\binom{52}{13} - \binom{48}{13}}.$$

Den sökta sannolikheten är

$$P(A^*|B) = 1 - P(A|B) = 1 - \frac{\binom{4}{1} \binom{48}{12}}{\binom{52}{13} - \binom{48}{13}} = \frac{5359}{14498} \approx 0.3696.$$

b) Givet att personen har hjärter ess är sannolikheten för inget ytterligare ess

$$P(A|\text{har hjärter ess}) = \frac{\binom{3}{0} \binom{48}{12}}{\binom{51}{12}}$$

och den sökta sannolikheten är

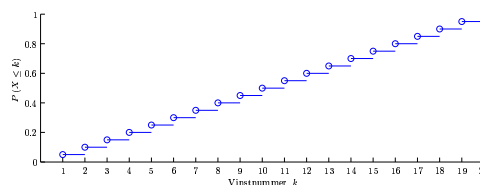
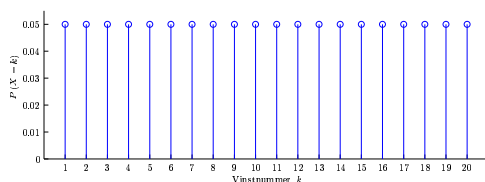
$$1 - P(A|\text{har hjärter ess}) = 1 - \frac{\binom{3}{0} \binom{48}{12}}{\binom{51}{12}} = \frac{11686}{20825} \approx 0.5612.$$

3.1 Låt X beteckna det nummer där lyckohjulet med n nummer stannar. De möjliga värdena på X är $\Omega_X = \{1, 2, \dots, n\}$. Ett rättvist hjul har samma sannolikhet för alla nummer, dvs $p_X(k) = P(X = k) = p$ beror ej av k . Eftersom

$$1 = \sum_{k \in \Omega_X} p_X(k) = \sum_{k=1}^n p = np$$

så är $p = 1/n$, dvs $P(X = k) = 1/n$ för $k \in \Omega_X$. Med $n = 20$ nummer så är

$$\begin{aligned} P(2 < X \leq 5) &= P(\{X = 3\} \cup \{X = 4\} \cup \{X = 5\}) = P(X = 3) + P(X = 4) + P(X = 5) \\ &= \frac{1}{n} + \frac{1}{n} + \frac{1}{n} = \frac{3}{n} = \frac{3}{20}. \end{aligned}$$



3.2 Lotterna i lotteriet fördelas enligt

Antal lotter	vinstbelopp
1	100
5	20
30	5
964	0
Totalt 1000	

Låt X beskriva vinstbeloppet av en på måfå vald lott. De möjliga värdena på X är $\Omega_X = \{0, 5, 20, 100\}$. Sannolikhetsfunktionen för X ges av

$$\begin{aligned} p_X(0) &= P(X = 0) = 964/1000 = 0.964 \\ p_X(5) &= P(X = 5) = 30/1000 = 0.030 \\ p_X(20) &= P(X = 20) = 5/1000 = 0.005 \\ p_X(100) &= P(X = 100) = 1/1000 = 0.001 \end{aligned}$$

och $p_X(x) = 0$ för övrigt.

3.3 De möjliga värdena för den stokastiska variabeln X ges av $\Omega_X = \{3, 4, 7, 8, 9\}$ där

$$p_X(3) = \frac{4}{12}, \quad p_X(4) = \frac{3}{12}, \quad p_X(7) = \frac{2}{12}, \quad p_X(8) = \frac{2}{12}.$$

a) Eftersom $\sum_{k \in \Omega_X} p_X(k) = 1$ så är

$$p_X(9) = 1 - [p_X(3) + p_X(4) + p_X(7) + p_X(8)] = \frac{1}{12}.$$

b) Eftersom fördelningsfunktionen $F_X(t) = P(X \leq t)$ fås

$$F_X(5) = P(X \leq 5) = \sum_{k \in \Omega_X: k \leq 5} p_X(k) = p_X(3) + p_X(4) = \frac{7}{12}.$$

c)

$$P(4 \leq X \leq 8) = \sum_{k \in \Omega_X: 4 \leq k \leq 8} p_X(k) = p_X(4) + p_X(7) + p_X(8) = \frac{7}{12}$$

och

$$P(X \geq 8) = \sum_{k \in \Omega_X: k \geq 8} p_X(k) = p_X(8) + p_X(9) = \frac{3}{12} = \frac{1}{4}.$$

3.4 Om X beskriver summan av två på måfå valda mynt så ges de möjliga värdena av $\Omega_X = \{2, 6\}$. Sannolikheten för att erhålla dessa värden skall nu bestämmas och uppgiften kan lösas på flera sätt.

Alternativ 1: Vi kan välja ut 2 mynt av tre på $\binom{3}{2} = 3$ sätt. Av dessa kan två enkronor väljas på 1 sätt så

$$P(X = 2) = P(\text{Välj två enkronor}) = \frac{1}{3}$$

och alltså $P(X = 6) = 1 - P(X \neq 6) = 1 - P(X = 2) = \frac{2}{3}$.

Alternativ 2: Låt A_1 och A_2 vara händelserna att det första respektive det andra valda myntet är en krona. Då är $P(A_1) = \frac{2}{3}$ och $P(A_2|A_1) = \frac{1}{2}$ så

$$P(X = 2) = P(\text{Välj två enkronor}) = P(A_1 \cap A_2) = P(A_2|A_1) P(A_1) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

och $P(X = 6) = 1 - P(X \neq 6) = 1 - P(X = 2) = \frac{2}{3}$.

Alternativ 3: De tre möjliga utfallen $\Omega = \{1_1 1_2, 1_1 5, 1_2 5\}$ lika stor sannolikhet.

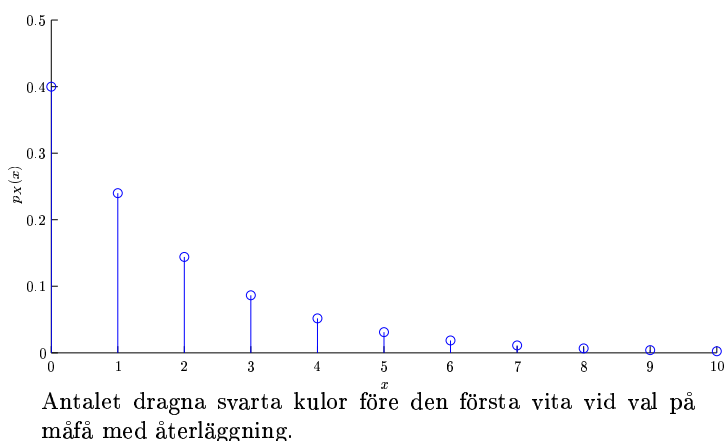
För de tre utfallen fås värdena $1 + 1 = 2$, $1 + 5 = 6$ och $1 + 5 = 6$. Den stokastiska variabeln X har således möjliga värden givna av $\Omega_X = \{2, 6\}$ där

$$p_X(2) = P(X = 2) = P(1_1 1_2) = \frac{1}{3} \quad p_X(6) = P(\{1_1 5\} \cup \{1_2 5\}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

3.5 I en urna med $N = 10$ kulor är $v = 4$ stycken vita och $s = 6$ stycken svarta. Vid dragning med återläggning låt X beskriva antalet dragna svarta kulor före den första vita kulan. Möjliga värden på X ges av $\Omega_X = \{0, 1, 2, 3, \dots\}$ och händelsen $X = k$, $k \in \Omega_X$, är händelsen att man dragit k svarta kulor följt av en vit. Alltså är

$$P(X = k) = \frac{\text{Antalet sätt att dra } k \text{ svarta följt av en vit}}{\text{Antalet sätt att dra } k + 1 \text{ kulor}} = \frac{s^k \cdot v}{N^{k+1}} = \left(\frac{s}{N}\right)^k \frac{v}{N} = (1-p)^k p,$$

för $k = 0, 1, 2, \dots$ där $p = v/N = 0.4$ andelen vita kulor i urnan. Detta är en geometrisk fördelning.



Nu är

$$P(X \geq 3) = \sum_{k \in \Omega_X: k \geq 3} p_X(k) = \sum_{k=3}^{\infty} (1-p)^k p = (1-p)^3 p \sum_{k=0}^{\infty} (1-p)^k = (1-p)^3 p \frac{1}{1-(1-p)} = (1-p)^3.$$

Detta kan även inses genom att händelsen $X \geq 3$ är händelsen att dra tre svarta kulor på rad. Med siffror är $P(X \geq 3) = (1-p)^3 = (0.6)^3 = 0.2160$.

3.6 I en urna med $N = 10$ kulor är $v = 4$ stycken vita och $s = 6$ stycken svarta. Vid dragning med återläggning låt Y beskriva antalet dragna kulor när den första vita kulan dras. Möjliga värden på Y ges av $\Omega_Y = \{1, 2, 3, \dots\}$ och händelsen $Y = k$, $k \in \Omega_Y$, är händelsen att man dragit $k - 1$ svarta kulor följt av en vit. Alltså är

$$P(Y = k) = \frac{\text{Antalet sätt att dra } k - 1 \text{ svarta följt av en vit}}{\text{Antalet sätt att dra } k \text{ kulor}} = \frac{s^{k-1} \cdot v}{N^k} = \left(\frac{s}{N}\right)^{k-1} \frac{v}{N} = (1-p)^{k-1} p,$$

för $k = 1, 2, 3, \dots$ där $p = v/N = 0.4$ andelen vita kulor i urnan. Detta är en för-första-gången-fördelning.

Notera att $Y = X + 1$ med X enligt uppgift 3.5.

Alternativ 1: Nu är

$$P(Y \geq 2) = \sum_{k \in \Omega_Y: k \geq 2} p_Y(k) = \sum_{k=2}^{\infty} (1-p)^{k-1} p = (1-p)p \sum_{k=0}^{\infty} (1-p)^k = (1-p)p \frac{1}{1-(1-p)} = 1-p.$$

Alternativ 2:

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - P(Y = 1) = 1 - (1-p)^0 p = 1 - p.$$

Händelsen $Y \geq 2$ är händelsen att dra en svart kula. Med siffror är $P(Y \geq 2) = 1 - p = 0.6$.

3.7 I en urna med $N = 10$ kulor är $v = 4$ stycken vita, det vill säga andelen vita kulor är $p = v/N = 0.40$.

a) Låt X beskriva antalet vita kulor vid dragning av $n = 3$ stycken med återläggning. Möjliga värden på X är $\Omega_X = \{0, 1, 2, 3\}$. Betrakta en sekvens av k vita och $n - k$ svarta kulor. En sådan sekvens har sannolikhet

$$\frac{\text{Antalet sätt att få en viss sekvens av } k \text{ vita och } n - k \text{ svarta}}{\text{Antalet sätt att dra } n \text{ kulor}} = \frac{v^k \cdot (N - v)^{n-k}}{N^n} = \left(\frac{v}{N}\right)^k \left(1 - \frac{v}{N}\right)^{n-k} = p^k (1 - p)^{n-k}.$$

Händelsen $X = k$ är händelsen att få någon av dessa sekvenser med k vita och $n - k$ svarta. Det finns $\binom{n}{k}$ sådana sekvenser så

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n.$$

Detta är binomialfördelningen.

Låt X beskriva antalet vita kulor vid dragning av $n = 3$ stycken utan återläggning. Möjliga värden på X är $\Omega_X = \{0, 1, 2, 3\}$.

$$P(X = k) = \frac{\text{Antalet sätt att få } k \text{ vita och } n - k \text{ svarta}}{\text{Antalet sätt att dra } n \text{ kulor}} = \frac{\binom{v}{k} \binom{N-v}{n-k}}{\binom{N}{n}} = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}} \quad k = 0, 1, 2, 3.$$

Detta är den hypergeometriska fördelningen.

b) Händelsen att få fler vita än svarta kulor är händelsen $\{X > n/2\} = \{X > 1.5\}$ som har sannolikhet

$$P(\text{Fler vita än svarta}) = \sum_{k \in \Omega_X : k > 1.5} p_X(k) = p_X(2) + p_X(3).$$

Vid dragningen med återläggning erhålls

$$p_X(2) + p_X(3) = \binom{3}{2} p^2 (1 - p)^1 + \binom{3}{3} p^3 (1 - p)^0 = p^2 (3 - 2p) = 0.3520$$

och vid dragning utan återläggning är sannolikheten

$$p_X(2) + p_X(3) = \frac{\binom{4}{2} \binom{6}{1}}{\binom{10}{3}} + \frac{\binom{4}{3} \binom{6}{0}}{\binom{10}{3}} = \frac{4 \cdot 3 \cdot 6 + 4 \cdot 1 \cdot 1}{3 \cdot 2 \cdot 1} = \frac{1}{3}.$$

3.8 De möjliga värdena på X är $\Omega_X = \{0, 1, 2, \dots\}$ där $P(X = k) = p^k$ för $k = 1, 2, 3, \dots$ och något $p > 0$. Nu är

$$\begin{aligned} 1 &= \sum_{k \in \Omega_X} p_X(k) = \sum_{k=0}^{\infty} P(X = k) = P(X = 0) + \sum_{k=1}^{\infty} P(X = k) \\ &= P(X = 0) + \sum_{k=1}^{\infty} p^k = P(X = 0) + p \sum_{k=0}^{\infty} p^k = P(X = 0) + p \frac{1}{1 - p}. \end{aligned}$$

så

$$P(X = 0) = 1 - \frac{p}{1 - p} = \frac{1 - 2p}{1 - p}.$$

Om uttrycken för $P(X = k)$, $k \in \Omega_X$, skall vara giltiga så skall $0 \leq P(X = k) \leq 1$ vilket ger

$$0 \leq \frac{1 - 2p}{1 - p} \leq 1$$

eller $0 < p \leq 1/2$.

3.9 Låt den stokastiska variabeln X har möjliga värden $\Omega_X = \{0, 1, 2, 3, \dots\}$ och sannolikhetsfunktion

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \dots,$$

och noll för övrigt. Vi söker sannolikheten $P(2 < X < 5)$. Alltså,

$$\begin{aligned} P(2 < X < 5) &= \sum_{\substack{k \in \Omega_X: \\ 2 < k < 5}} p_X(k) = p_X(3) + p_X(4) = \frac{\mu^3}{3!} e^{-\mu} + \frac{\mu^4}{4!} e^{-\mu} = \frac{(4 + \mu)\mu^3}{24} e^{-\mu} = \frac{128}{6} e^{-4} \\ &\approx 0.3907. \end{aligned}$$

3.10 Låt A_k stå för händelsen att terminal k används, $k = 1, 2, 3$. Om X beskriver antalet terminaler som används så ges de möjliga värdena för X av $\Omega_X = \{0, 1, 2, 3\}$. Vi erhåller att

$$P(X = 0) = P(A_1^* \cap A_2^* \cap A_3^*) = \{\text{oberoende}\} = P(A_1^*) P(A_2^*) P(A_3^*) = \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{24}.$$

Vi ser också att

$$P(X = 3) = P(A_1 \cap A_2 \cap A_3) = \{\text{oberoende}\} = P(A_1) P(A_2) P(A_3) = \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{6}{24}.$$

Händelsen $\{X = 2\}$ är händelsen $\{A_1 \cap A_2 \cap A_3^*\} \cup \{A_1 \cap A_2^* \cap A_3\} \cup \{A_1^* \cap A_2 \cap A_3\}$. Vi erhåller

$$\begin{aligned} P(X = 2) &= P((A_1 \cap A_2 \cap A_3^*) \cup (A_1 \cap A_2^* \cap A_3) \cup (A_1^* \cap A_2 \cap A_3)) \\ &= P(A_1 \cap A_2 \cap A_3^*) + P(A_1 \cap A_2^* \cap A_3) + P(A_1^* \cap A_2 \cap A_3) \\ &= P(A_1) P(A_2) P(A_3^*) + P(A_1) P(A_2^*) P(A_3) + P(A_1^*) P(A_2) P(A_3) \\ &= \frac{11}{24}. \end{aligned}$$

Den sista sannolikheten $P(X = 1)$ fås enklast som $P(X = 1) = 1 - P(X \neq 1) = 1 - (P(X = 0) + P(X = 2) + P(X = 3))$. Vi har alltså

$$p_X(k) = \begin{cases} \frac{1}{24} & \text{för } k = 0, \\ \frac{6}{24} & \text{för } k = 1, \\ \frac{11}{24} & \text{för } k = 2, \\ \frac{6}{24} & \text{för } k = 3, \\ 0 & \text{för övrigt.} \end{cases}$$

3.11 Enligt uppgift är

$$\frac{1}{2} = P(X = 0) = \frac{\mu^0}{0!} e^{-\mu} = e^{-\mu}$$

vilket ger att $\mu = -\ln(1/2) = \ln(2)$. Sålunda har vi att

$$P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - \left(\frac{1}{2} + \frac{\mu^1}{1!} e^{-\mu} \right) = \frac{1 - \ln(2)}{2} \approx 0.1534.$$

3.12 De möjliga värdena på X ges av intervallet $\Omega_X = [0, 10)$. Intervallet innehåller ett överuppräknligt antal värden och dessa kan inte alla tillskrivas positiva sannolikheter.

Ett rimligt utseende på fördelningsfunktionen är

$$F_X(t) = P(X \leq t) = \frac{t}{10}$$

för $0 \leq t < 10$, eftersom händelsen $\{X \leq t\}$, att få vänta högst t minuter, innebär att mannen kom till hållplatsen under ett tidsintervall av längd t minuter före nästa tåg. Med 10 minuter mellan tågen är sannolikheten att träffa ett intervall av längd t , $t/10$, om tidpunkten väljes på måfå.

Ur

$$P(a < X \leq b) = F_X(b) - F_X(a) \quad \text{och} \quad P(a < X \leq b) = \int_a^b f_X(x) dx$$

fås att $F_X(t)$ är primitiv funktion till $f_X(t)$ och

$$f_X(t) = \frac{d}{dt} F_X(t) = \frac{1}{10}, \quad 0 \leq t \leq 10,$$

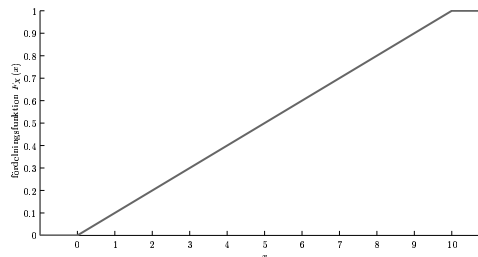
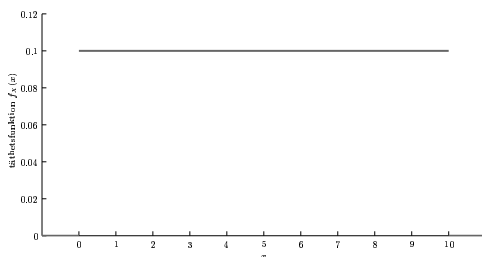
likformig fördelning på intervallet $[0, 10]$.

b) Alternativ 1:

$$P(3.5 < X \leq 7) = \int_{\substack{x \in \Omega_X \\ 3.5 < x \leq 7}} f_X(x) dx = \int_{3.5}^7 \frac{1}{10} dx = \frac{7 - 3.5}{10} = 0.35.$$

Alternativ 2:

$$P(3.5 < X \leq 7) = F_X(7) - F_X(3.5) = \frac{7}{10} - \frac{3.5}{10} = \frac{7 - 3.5}{10} = 0.35.$$



3.13 En kontinuerlig stokastisk variabel med täthetsfunktion $f_X(x)$ uppfyller att

$$1 = \int_{-\infty}^{\infty} f_X(x) dx.$$

Om $f_X(x) = f(x)$ så är

$$1 = \int_{-\infty}^{\infty} f_X(x) dx = \int_0^6 cx^2 dx = c \left[\frac{x^3}{3} \right]_0^6 = 72c$$

det vill säga $c = 1/72$.

3.14 En kontinuerlig stokastisk variabel med täthetsfunktion $f_X(x)$ uppfyller att

$$1 = \int_{-\infty}^{\infty} f_X(x) dx.$$

Om $f_X(x) = f(x)$ så är

$$1 = \int_{-\infty}^{\infty} f_X(x) dx = \int_{-1}^1 \frac{c}{\sqrt{x+1}} dx = 2c [\sqrt{x+1}]_{-1}^1 = 2\sqrt{2}c$$

det vill säga $c = 1/\sqrt{8}$. Med detta värde är

$$P(X > 0) = \int_0^{\infty} f_X(x) dx = \int_0^1 \frac{c}{\sqrt{x+1}} dx = 2c [\sqrt{x+1}]_0^1 = 2c(\sqrt{2} - 1) = 1 - \frac{1}{\sqrt{2}} \approx 0.2929$$

3.15 Den stokastiska variabeln X har fördelningsfunktion

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{för } x < 1 \\ 1 - 1/x^2 & \text{för } x \geq 1. \end{cases}$$

Vi söker medianen $\tilde{x}_{0.50}$, den punkt sådan att

$$F_X(\tilde{x}_{0.50}) = \frac{1}{2}.$$

Det innebär att

$$\frac{1}{2} = F_X(\tilde{x}_{0.50}) = 1 - \frac{1}{\tilde{x}_{0.50}^2}$$

vilket har lösningarna $\tilde{x}_{0.50} = -\sqrt{2}$ och $\tilde{x}_{0.50} = \sqrt{2}$. Kravet $\tilde{x}_{0.50} > 1$ ger lösningen $\tilde{x}_{0.50} = \sqrt{2}$.

3.16 En stokastisk variabel X med täthetsfunktion

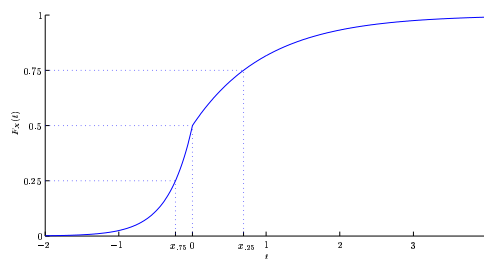
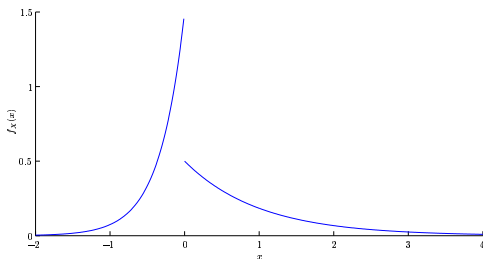
$$f_X(x) = \begin{cases} \frac{3}{2}e^{3x} & \text{om } x < 0 \\ \frac{1}{2}e^{-x} & \text{om } x \geq 0 \end{cases}$$

har fördelningsfunktion $F_X(t) = P(X \leq t)$ som för $t \leq 0$ är

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \int_{-\infty}^t \frac{3}{2}e^{3x} dx = \frac{3}{2} \left[e^{3x} \cdot \frac{1}{3} \right]_{-\infty}^t = \frac{1}{2}e^{3t}.$$

För $t > 0$ är

$$\begin{aligned} F_X(t) &= \int_{-\infty}^t f_X(x) dx = \int_{-\infty}^0 f_X(x) dx + \int_0^t f_X(x) dx = F_X(0) + \int_0^t \frac{1}{2}e^{-x} dx \\ &= \frac{1}{2} + \frac{1}{2} [e^{-x}(-1)]_0^t = 1 - \frac{1}{2}e^{-t}. \end{aligned}$$



En α -kvantil x_α är lösningen till $F_X(x_\alpha) = 1 - \alpha$. Eftersom $F_X(0) = \frac{1}{2}$ är medianen $\tilde{x}_{0.50} = 0$ och $x_{0.75} \leq 0$ och $x_{0.25} \geq 0$ varför $x_{0.75}$ bestäms ur

$$\frac{1}{4} = F_X(x_{0.75}) = \frac{1}{2}e^{-3x_{0.75}} \Rightarrow x_{0.75} = -\frac{1}{3} \ln(2) \approx -0.2310$$

och $x_{0.25}$ bestäms ur

$$\frac{3}{4} = F_X(x_{0.25}) = 1 - \frac{1}{2}e^{-x_{0.25}} \Rightarrow x_{0.25} = \ln(2) \approx 0.6931.$$

3.17 Låt X beskriva tiden från butikens öppnande till första kunden kommer. Då är

a) $P(X \leq 3) = F_X(3) = 1 - e^{-0.4 \cdot 3} = 1 - e^{-1.2} \approx 0.6988.$

b) $P(X \geq 4) = P(X > 4) = 1 - F_X(4) = 1 - (1 - e^{-0.4 \cdot 4}) = e^{-1.6} \approx 0.2019.$

c) $P(3 \leq X \leq 4) = P(3 < X \leq 4) = F_X(4) - F_X(3) = (1 - e^{-0.4 \cdot 4}) - (1 - e^{-0.4 \cdot 3}) = e^{-1.2} - e^{-1.6} \approx 0.0993.$

d) $P(\{X \leq 3\} \cup \{X \geq 4\}) = P(\{X \leq 3\} \cup \{X > 4\}) = 1 - P(3 < X \leq 4) = 1 - (e^{-1.2} - e^{-1.6}) \approx 0.9007.$

e) Fördelningsfunktionen är kontinuerlig så $P(X = x) = 0$ för alla x .

3.18 En stokastisk variabel X med täthetsfunktion

$$f_X(x) = \begin{cases} 2xe^{-x^2} & \text{om } x > 0 \\ 0 & \text{om } x \leq 0 \end{cases}$$

har fördelningsfunktion $F_X(t) = P(X \leq t)$ som för $t > 0$ är

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \int_0^t 2xe^{-x^2} dx = \{u = x^2\} = \int_0^{t^2} e^{-u} du = [e^{-u}(-1)]_0^{t^2} = 1 - e^{-t^2}$$

och $F_X(t) = 0$ för $t \leq 0$.

En α -kvantil x_α är lösningen till $F_X(x_\alpha) = 1 - \alpha$ vilket ger att för $0 < \alpha < 1$ är

$$1 - \alpha = 1 - e^{-x_\alpha^2}$$

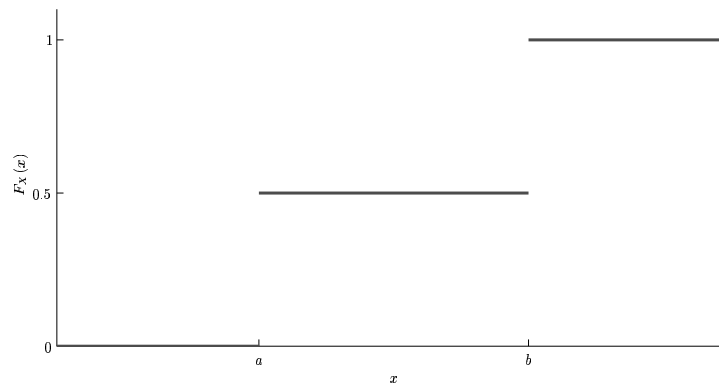
vilket har lösningarna $x_\alpha = \pm \sqrt{-\ln(\alpha)}$. Eftersom $F_X(0) = 0$ är $x_\alpha > 0$ och de negativa rötterna är inga lösningar. Alltså, α -kvantilen är $x_\alpha = \sqrt{-\ln(\alpha)}$.

Numeriska värden:

α	x_α
0.5	0.8326
0.1	1.5174
0.01	2.1460.

3.19 Fördelningsfunktionen för tvåpunktsfördelningen med massan $1/2$ i punkterna a och b där $a < b$ har fördelningsfunktion:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{då } x < a \\ \frac{1}{2} & \text{då } a \leq x < b \\ 1 & \text{då } x \geq b. \end{cases}$$



Fördelningsfunktionen $F_X(x)$ för tvåpunktsfördelningen med massan $1/2$ i punkterna a och b .

3.20 Låt X beskriva längden av ett telefonsamtal. Givet är att för $t \geq 0$ är

$$e^{-\lambda t} = P(\text{Samtal längre än } t) = P(X > t)$$

för någon konstant $\lambda > 0$. Alltså är fördelningsfunktionen för X :

$$F_X(t) = P(X \leq t) = 1 - P(X > t) = 1 - e^{-\lambda t}$$

för $t \geq 0$. Således har X täthetsfunktion

$$f_X(t) = \frac{d}{dt}F_X(t) = 0 - e^{-\lambda t}(-\lambda) = \lambda e^{-\lambda t},$$

för $t \geq 0$. Sannolikheten $P(1 < X \leq 10)$ bestäms på två alternativa sätt:

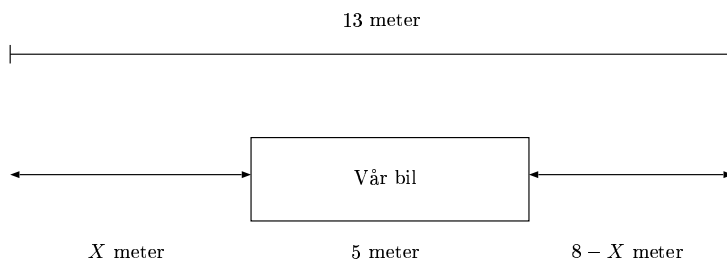
Alternativ 1: Med $\lambda = 2/3$ fås

$$P(1 < X \leq 10) = \int_1^{10} f_X(x) dx = \int_1^{10} \lambda e^{-\lambda x} dx = \left[\lambda e^{-\lambda x} \frac{-1}{\lambda} \right]_1^{10} = e^{-\lambda} - e^{-10\lambda} \approx 0.5121.$$

Alternativ 2: Med $\lambda = 2/3$ fås

$$P(1 < X \leq 10) = F_X(10) - F_X(1) = (1 - e^{-\lambda \cdot 10}) - (1 - e^{-\lambda \cdot 1}) = e^{-\lambda} - e^{-10\lambda} \approx 0.5121.$$

3.21 Låt X beskriva avståndet mellan parkeringsfickans början och bilen. Om vi parkerar bilen på måfå i fickan ansätter vi modellen att X är likformigt fördelad på intervallet 0 till $13 - 5 = 8$ meter, dvs. $f_X(x) = 1/8$ då $0 \leq x \leq 8$.



En annan bil får plats om vår bil står tidigt ($X < 3$) eller sent ($X > 5$) i fickan. Med siffror

$$P(\{\text{annan bil får plats}\}) = P(\{X < 3\} \cup \{X > 5\}) = \int_0^3 f_X(x) dx + \int_5^8 f_X(x) dx = \frac{3}{4}.$$

3.22 Låt X beskriva livslängden för en transistor. Modell: X är exponentialfördelad med parameter $\lambda = 10^{-4}$, det vill säga för $x \geq 0$ är

$$f_X(x) = \lambda e^{-\lambda x} = \frac{1}{10000} e^{-x/10000},$$

$f_X(x) = 0$ om $x < 0$.

a) Vi erhåller

$$P(X < 6000) = \int_{-\infty}^{6000} f_X(x) dx = \int_0^{6000} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^{6000} = 1 - e^{-0.6} \approx 0.4512.$$

b) Låt A_k vara händelsen att transistor k upphör att fungera inom 6000 timmar, $k=1,2,3,4,5$. Händelsen

$$\{\text{minst en upphör att fungera}\} = A_1 \cup A_2 \cup \dots \cup A_5 = (A_1^* \cap A_2^* \cap \dots \cap A_5^*)^*$$

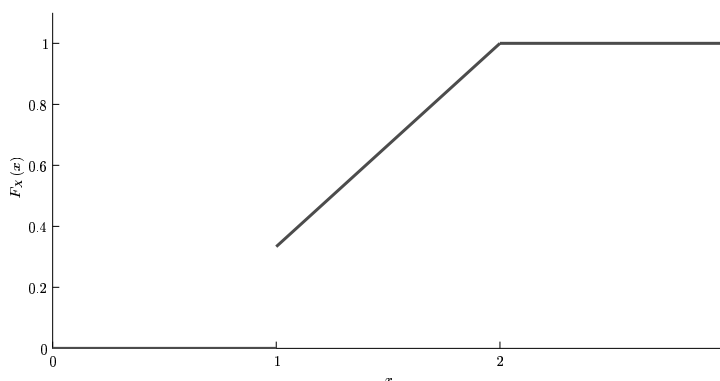
det vill säga motsatsen till att någon upphör att fungera inom 6000 timmar är att alla fungerar minst 6000 timmar. Härav får vi att

$$P(\text{minst en upphör fungera inom 6000 timmar}) = 1 - P(\text{alla fungerar minst 6000 timmar})$$

och med hjälp av oberoendet mellan transistorerna fås

$$\begin{aligned} P(\text{minst en upphör fungera inom 6000 timmar}) &= 1 - P(A_1^* \cap A_2^* \cap \dots \cap A_5^*) \\ &= 1 - P(A_1^*) P(A_2^*) \dots P(A_5^*) = 1 - (1 - 0.4512)^5 \approx 0.9502. \end{aligned}$$

3.23



Fördelningsfunktionen $F_X(x)$ med diskontinuitet i $x = 1$.

a) $P(X \leq 5/3) = F_X(5/3) = \frac{1}{3} + \frac{2}{3} \left(\frac{5}{3} - 1\right) = \frac{7}{9}$.

b) $P(X > 3/2) = 1 - P(X \leq 3/2) = 1 - F_X(3/2) = 1 - \left(\frac{1}{3} + \frac{2}{3} \left(\frac{3}{2} - 1\right)\right) = \frac{1}{3}$.

c) $P(4/3 < X \leq 5/3) = F_X(5/3) - F_X(4/3) = \left(\frac{1}{3} + \frac{2}{3} \left(\frac{5}{3} - 1\right)\right) - \left(\frac{1}{3} + \frac{2}{3} \left(\frac{4}{3} - 1\right)\right) = \frac{2}{9}$.

d)

$$\begin{aligned} P(X = 1) &= P(X \leq 1) - P(X < 1) = F_X(1) - \lim_{h \rightarrow 1^-} P(X \leq h) \\ &= F_X(1) - \lim_{h \rightarrow 1^-} \underbrace{F_X(h)}_{=0} = F_X(1) = \frac{1}{3}. \end{aligned}$$

3.24 Intensiteten $\lambda_X(x)$ för en (positiv) kontinuerlig stokastisk variabel X ges av relationen

$$\lambda_X(x) = \frac{f_X(x)}{1 - F_X(x)}$$

där $f_X(x)$ är täthetsfunktionen och $F_X(x)$ är fördelningsfunktionen för den stokastiska variabeln. Från fördelningsfunktionen

$$F_X(x) = 1 - \frac{1}{(1+x)^c}$$

bestäms täthetsfunktionen

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{c}{(1+x)^{c+1}}$$

vilket ger intensiteten

$$\lambda_X(x) = \frac{c/(1+x)^{c+1}}{1/(1+x)^c} = \frac{c}{1+x}$$

för $x \geq 0$.

3.25 Överlevnadsfunktionen $R(t) = P(X > t)$ bestäms ur intensiteten $\lambda_X(x)$ för en (positiv) kontinuerlig stokastisk variabel genom relationen

$$R(t) = \exp \left\{ - \int_0^t \lambda_X(x) dx \right\}.$$

Här är för $t > 0$

$$\int_0^t \lambda_X(x) dx = \int_0^t e^x dx = [e^x]_0^t = e^t - 1$$

så

$$R(t) = \exp\{1 - e^t\}$$

och fördelningsfunktionen blir

$$F_X(t) = P(X \leq t) = 1 - P(X > t) = 1 - R(t) = 1 - e^{1-e^t}$$

för $t \geq 0$.

3.26 Den diskreta stokastiska variabeln X är likformigt fördelad över de möjliga värdena $\Omega_X = \{1, 2, \dots, 6\}$.

För funktionen $y(x) = 2x$ gäller att

$$\begin{array}{l} x: \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\ y(x): \quad 2 \quad 4 \quad 6 \quad 8 \quad 10 \quad 12 \end{array}$$

så den stokastiska variabeln $Y = y(X)$ har möjliga värden givna av $\Omega_Y = \{2, 4, 6, 8, 10, 12\}$ där

$$\begin{array}{ll} P(Y = 2) = P(X = 1) = \frac{1}{6} & P(Y = 8) = P(X = 4) = \frac{1}{6} \\ P(Y = 4) = P(X = 2) = \frac{1}{6} & P(Y = 10) = P(X = 5) = \frac{1}{6} \\ P(Y = 6) = P(X = 3) = \frac{1}{6} & P(Y = 12) = P(X = 6) = \frac{1}{6}. \end{array}$$

3.27 Den diskreta stokastiska variabeln X är likformigt fördelad över de möjliga värdena $\Omega_X = \{0, 1, 2, \dots, 9\}$.

För funktionen $y(x) = |x - 3|$ gäller att

$$\begin{array}{l} x: \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \\ y(x): \quad 3 \quad 2 \quad 1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \end{array}$$

så den stokastiska variabeln $Y = y(X)$ har möjliga värden givna av $\Omega_Y = \{0, 1, \dots, 6\}$ där

$$\begin{array}{ll} P(Y = 0) = P(X = 3) = \frac{1}{10} & P(Y = 4) = P(X = 7) = \frac{1}{10} \\ P(Y = 1) = P(X = 2) + P(X = 4) = \frac{2}{20} & P(Y = 5) = P(X = 8) = \frac{1}{10} \\ P(Y = 2) = P(X = 1) + P(X = 5) = \frac{2}{20} & P(Y = 6) = P(X = 9) = \frac{1}{10} \\ P(Y = 3) = P(X = 0) + P(X = 6) = \frac{2}{20} & \end{array}$$

3.28 Låt X vara likformigt fördelad på intervallet $[-1, 1]$, det vill säga

$$f_X(x) = \frac{1}{2} \quad -1 \leq x \leq 1$$

och $f_X(x) = 0$ för övrigt. Då har X fördelningsfunktion

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx = \begin{cases} 0 & \text{om } t < -1 \\ \frac{t+1}{2} & \text{om } -1 \leq t < 1 \\ 1 & \text{om } t \geq 1. \end{cases}$$

Med $Y = (X+1)/2$ så ges de möjliga värdena på Y av $\Omega_Y = [0, 1]$ och för $t \in \Omega_Y$ har Y fördelningsfunktion

$$F_Y(t) = P(Y \leq t) = P\left(\frac{X+1}{2} \leq t\right) = P(X \leq 2t-1) = F_X(2t-1) = \frac{(2t-1)+1}{2} = t$$

det vill säga $f_Y(t) = 1$ för $0 \leq t \leq 1$ och Y är likformigt fördelad på intervallet $[0, 1]$.

3.29 Låt X vara en positiv kontinuerlig stokastisk variabel med täthetsfunktion $f_X(x)$ och fördelningsfunktion $F_X(x)$. Med $Y = 1/X$ så är Y en positiv stokastisk variabel med fördelningsfunktion

$$F_Y(x) = P(Y \leq x) = P\left(\frac{1}{X} \leq x\right) = P\left(X \geq \frac{1}{x}\right) = 1 - P\left(X < \frac{1}{x}\right) = 1 - F_X\left(\frac{1}{x}\right)$$

för $x > 0$. Derivering ger med kedjeregeln

$$f_Y(x) = \frac{d}{dx} F_Y(x) = -f_X\left(\frac{1}{x}\right) \cdot \frac{-1}{x^2} = \frac{1}{x^2} f_X\left(\frac{1}{x}\right).$$

Med given täthet $f_X(x)$ erhålles

$$f_Y(x) = \frac{1}{x^2} \cdot \frac{1}{\pi} \cdot \frac{2}{1+(1/x)^2} = \frac{1}{\pi} \cdot \frac{2}{1+x^2} = f_X(x)$$

det vill säga X och Y har samma fördelning (är likafördelade).

3.30 Låt X beskriva den tid som fru Svensson parkerar. Enligt modellen är X likformigt fördelad på intervallet $[20, 60]$, det vill säga

$$f_X(x) = \frac{1}{40} \quad 20 \leq x \leq 60.$$

Med Y som den avgift fru Svensson betalar i parkeringsavgift ges de möjliga värdena på Y av $\Omega_Y = \{14, 16, 18\}$. Vidare så är

$$\begin{aligned} p_Y(14) &= P(15 \leq X < 30) = \int_{15}^{30} f_X(x) dx = \int_{20}^{30} f_X(x) dx = \frac{1}{4}. \\ p_Y(16) &= P(30 \leq X < 45) = \int_{30}^{45} f_X(x) dx = \frac{3}{8} \\ p_Y(18) &= P(45 \leq X < 60) = \int_{45}^{60} f_X(x) dx = \frac{3}{8}. \end{aligned}$$

3.31 Låt X vara exponentialfördelad, det vill säga X har täthetsfunktion

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

för någon konstant $\lambda > 0$. Låt $Y = \lfloor X \rfloor$ vara heltalsdelen av X . De möjliga värdena på Y ges av $\Omega_Y = \{0, 1, 2, 3, \dots\}$ och för $k \in \Omega_Y$ är

$$\begin{aligned} P(Y = k) &= P(k \leq X < k+1) = \int_k^{k+1} f_X(x) dx = \int_k^{k+1} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_k^{k+1} \\ &= e^{-\lambda k} - e^{-\lambda(k+1)} = e^{-\lambda k} [1 - e^{-\lambda}] = p^k (1 - p) \end{aligned}$$

där $p = e^{-\lambda}$. Alltså är $\lfloor X \rfloor$ geometriskt fördelad med parameter $e^{-\lambda}$.

3.32 Låt X beskriva positionen för en jordbävningens epicentrum i förkastningssprickan. Enligt modellen är X likformigt fördelad på intervallet $[-a, a]$, det vill säga

$$f_X(x) = \frac{1}{2a}, \quad -a \leq x \leq a.$$

Med Y som avståndet mellan damm och epicentrum fås med Pythagoras sats

$$Y = \sqrt{X^2 + d^2}.$$

De möjliga värdena på Y är $\Omega_Y = [d, \sqrt{a^2 + d^2}]$. För ett $t \in \Omega_Y$ så är

$$\begin{aligned} F_Y(t) &= P(Y \leq t) = P(\sqrt{X^2 + d^2} \leq t) = P(X^2 \leq t^2 - d^2) = P(|X| \leq \sqrt{t^2 - d^2}) \\ &= P(-\sqrt{t^2 - d^2} \leq X \leq \sqrt{t^2 - d^2}) = \int_{-\sqrt{t^2 - d^2}}^{\sqrt{t^2 - d^2}} f_X(x) dx = \frac{\sqrt{t^2 - d^2}}{a}. \end{aligned}$$

Täthetsfunktionen fås genom derivering

$$f_Y(t) = \frac{d}{dt} F_Y(t) = \frac{d}{dt} \left[\frac{\sqrt{t^2 - d^2}}{a} \right] = \frac{t}{a\sqrt{t^2 - d^2}}, \quad \text{för } d \leq t \leq \sqrt{a^2 + d^2}.$$

4.1 Sannolikhetsfunktionen för (X, Y) ges av tabellen

$p_{X,Y}(j, k)$		k				$p_X(j)$
		1	2	3	4	
j	0	0.11	0.09	0.07	0.01	0.28
	1	0.07	0.12	0.12	0.02	0.33
	2	0.02	0.05	0.17	0.05	0.29
	3	0.00	0.02	0.04	0.02	0.08
	4	0.00	0.00	0.01	0.01	0.02
		$p_Y(k)$	0.20	0.28	0.41	0.11

där marginalfördelningarna $p_X(j)$ och $p_Y(k)$ bestämts ur

$$p_X(j) = \sum_{k:(j,k) \in \Omega_{X,Y}} p_{X,Y}(j, k) \quad p_Y(k) = \sum_{j:(j,k) \in \Omega_{X,Y}} p_{X,Y}(j, k).$$

a) Ur den simultana fördelningen får man att

$$P(X \leq 1, Y \leq 3) = \sum_{\substack{(j,k) \in \Omega_{X,Y}: \\ j \leq 1, k \leq 3}} p_{X,Y}(j, k) = \sum_{j=0}^1 \sum_{k=1}^3 p_{X,Y}(j, k) = .11 + .09 + .07 + .07 + .12 + .12 = 0.58.$$

b) En familj är trångbodd om $(X + 2)/Y > 2$ och

$$\begin{aligned} P(X + 2 > 2Y) &= \sum_{\substack{(j,k) \in \Omega_{X,Y}: \\ (j+2) > 2k}} p_{X,Y}(j, k) = \sum_{j=0}^4 \sum_{k=1}^{j/2} p_{X,Y}(j, k) \\ &= p_{X,Y}(1, 1) + p_{X,Y}(2, 1) + p_{X,Y}(3, 1) + p_{X,Y}(3, 2) + p_{X,Y}(4, 1) + p_{X,Y}(4, 2) = 0.11 \end{aligned}$$

4.2 a) Med oberoendet mellan X och Y erhålls

$$P(X = 3, Y = 6) = P(X = 3)P(Y = 6) = p_X(3)p_Y(6) = 0.20 \cdot 0.30 = 0.06.$$

b) Vidare, med oberoendet så är $P(X \leq 3, Y \leq 6) = P(X \leq 3)P(Y \leq 6)$ där

$$P(X \leq 3) = P(X = 1) + P(X = 3) = 0.10 + 0.20 = 0.30$$

och

$$P(Y \leq 6) = P(Y = 2) + P(Y = 4) + P(Y = 6) = 0.10 + 0.20 + 0.30 = 0.60.$$

Sålunda,

$$P(X \leq 3, Y \leq 6) = P(X \leq 3)P(Y \leq 6) = 0.30 \cdot 0.60 = 0.18.$$

4.3 För en tvådimensionell stokastisk variabel (X, Y) gäller att

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx = \int_1^{\infty} \int_0^{\infty} \frac{c}{x} \cdot e^{-x^2 y} dy dx = \int_1^{\infty} \frac{c}{x} \left[e^{-x^2 y} \frac{-1}{x^2} \right]_0^{\infty} dx \\ &= \int_1^{\infty} \frac{c}{x} \cdot \frac{1}{x^2} dx = \left[\frac{-c}{2x^2} \right]_1^{\infty} = \frac{c}{2}, \end{aligned}$$

det vill säga $c = 2$.

4.4 Den tvådimensionella stokastiska variabeln (X, Y) har täthetsfunktion

$$f_{X,Y}(x, y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}.$$

Man bestämmer $F_{X,Y}(s, t) = P(X \leq s, Y \leq t)$ ur

$$\begin{aligned} F_{X,Y}(s, t) &= \int_{-\infty}^s \int_{-\infty}^t f_{X,Y}(x, y) dy dx = \int_{-\infty}^s \int_{-\infty}^t \frac{1}{\pi^2(1+x^2)(1+y^2)} dy dx \\ &= \frac{1}{\pi^2} \int_{-\infty}^s \frac{1}{1+x^2} \int_{-\infty}^t \frac{1}{1+y^2} dy dx = \frac{1}{\pi^2} \int_{-\infty}^s \frac{1}{1+x^2} [\tan^{-1}(y)]_{-\infty}^t dx \\ &= \frac{1}{\pi^2} \int_{-\infty}^s \frac{1}{1+x^2} \left[\tan^{-1}(t) + \frac{\pi}{2} \right] dx = \frac{\tan^{-1}(t) + \pi/2}{\pi^2} \int_{-\infty}^s \frac{1}{1+x^2} dx \\ &= \frac{\tan^{-1}(t) + \pi/2}{\pi^2} [\tan^{-1}(x)]_{-\infty}^s = \frac{(\tan^{-1}(s) + \pi/2)(\tan^{-1}(t) + \pi/2)}{\pi^2}. \end{aligned}$$

Ur $F_{X,Y}(x, y)$ kan de marginella fördelningsfunktionerna bestämmas

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \leq \infty) = F_{X,Y}(x, \infty) = \frac{\tan^{-1}(x) + \pi/2}{\pi}.$$

På samma sätt erhålls $F_Y(y) = P(Y \leq y) = P(X \leq \infty, Y \leq y) = F_{X,Y}(\infty, y) = \frac{\tan^{-1}(y) + \pi/2}{\pi}$. Deriveras de marginella fördelningsfunktionerna erhålls de marginella täthetsfunktionerna:

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} \frac{\tan^{-1}(x) + \pi/2}{\pi} = \frac{1}{\pi(1+x^2)}$$

och $f_Y(y) = \frac{1}{\pi(1+y^2)}$. Notera att $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ för alla (x, y) varför de stokastiska variablerna X och Y är oberoende.

4.5 Ur tabellen har man att

$$P(Y = k_1) = P(X = j_1, Y = k_1) + P(X = j_2, Y = k_1) + P(X = j_3, Y = k_1) = .03 + .04 + .03 = 0.10.$$

Med oberoendet fås $P(X = j, Y = k_1) = P(X = j) P(Y = k_1)$, $j = j_1, j_2, j_3$, vilket ger

$$P(X = j_1) = \frac{P(X = j_1, Y = k_1)}{P(Y = k_1)} = \frac{0.03}{0.10} = 0.30$$

$$P(X = j_2) = \frac{P(X = j_2, Y = k_1)}{P(Y = k_1)} = \frac{0.04}{0.10} = 0.40$$

$$P(X = j_3) = \frac{P(X = j_3, Y = k_1)}{P(Y = k_1)} = \frac{0.03}{0.10} = 0.30.$$

På samma sätt är

$$P(Y = k_2) = \frac{P(X = j_1, Y = k_2)}{P(X = j_1)} = \frac{0.15}{0.30} = 0.50$$

och slutligen

$$P(Y = k_3) = 1 - P(Y \neq k_3) = 1 - (P(Y = k_2) + P(Y = k_1)) = 1 - 0.10 - 0.50 = 0.40.$$

Ur marginalfördelningarna bestäms $P(X = j, Y = k) = P(X = j) P(Y = k)$:

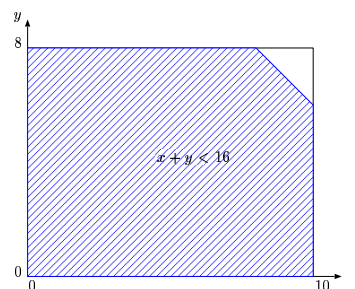
	k_1	k_2	k_3
j_1	0.03	0.15	0.12
j_2	0.04	0.20	0.16
j_3	0.03	0.15	0.12

4.6

Då X har $f_X(x) = 1/10$ för $0 \leq x \leq 10$ och Y har $f_Y(y) = 1/8$ för $0 \leq y \leq 8$ och X och Y är oberoende så är

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{10} \cdot \frac{1}{8} = \frac{1}{80}$$

för $(x, y) \in [0, 10] \times [0, 8]$.



Paret av väntetider (X, Y) för de två bussarna är likformigt fördelat över rektangeln $[0, 10] \times [0, 8]$.

Alternativ 1: Enligt definitionen

$$\begin{aligned} P(X + Y \geq 16) &= \iint_{(x,y): x+y \geq 16} f_{X,Y}(x, y) dy dx = \int_8^{10} \int_{16-x}^8 f_{X,Y}(x, y) dy dx \\ &= \int_8^{10} \int_{16-x}^8 \frac{1}{80} dy dx = \int_8^{10} \frac{x-8}{80} dx = \frac{1}{80} \left[\frac{x^2}{2} - 8x \right]_8^{10} = \frac{1}{40}. \end{aligned}$$

Alternativ 2: Ur figuren. Under likformig fördelning att få ett utfall i det omarkerade området, $x + y \geq 16$, ges som förhållandet mellan areorna. Rektangeln har area $10 \cdot 8 = 80$. Triangeln med hörn i $(10, 6)$, $(10, 8)$ och $(8, 8)$ har area $2 \cdot 2/2 = 2$. Sannolikheten är alltså

$$P(X + Y \geq 16) = \frac{2}{80} = \frac{1}{40}.$$

4.7 Då X har $f_X(x) = 1/4$ för $0 \leq x \leq 4$ och Y har $f_Y(y) = 1/6$ för $0 \leq y \leq 6$ och X och Y är oberoende så är

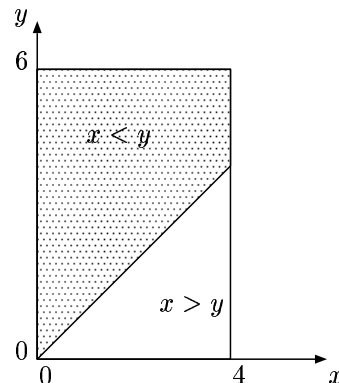
$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24}$$

för $(x,y) \in [0,4] \times [0,6]$. Bestäm $P(X < Y)$.

Alternativ 1: Enligt definitionen

$$\begin{aligned} P(X < Y) &= \iint_{x < y} f_{X,Y}(x,y) dy dx = \int_0^4 \int_x^6 f_{X,Y}(x,y) dy dx = \int_0^4 \int_x^6 \frac{1}{24} dy dx \\ &= \int_0^4 \frac{6-x}{24} dx = \frac{1}{24} \left[6x - \frac{x^2}{2} \right]_0^4 = \frac{2}{3}. \end{aligned}$$

Alternativ 2: Ur figuren. Att få ett utfall i det markerade området under likformig fördelning ges som förhållandet mellan areorna. Rektangeln har area $4 \cdot 6 = 24$. Triangeln med hörn i $(0,0)$, $(4,4)$ och $(4,0)$ har area $4 \cdot 4/2 = 8$. Sannolikheten är alltså $P(X < Y) = \frac{24-8}{24} = \frac{2}{3}$.

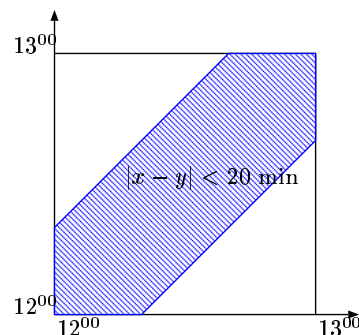


4.8

Med enheten timmar, är $f_X(x) = 1$, $0 \leq x \leq 1$, $f_Y(y) = 1$, $0 \leq y \leq 1$, och med oberoendet $f_{X,Y}(x,y) = f_X(x)f_Y(y) = 1 \cdot 1 = 1$ för $(x,y) \in [0,1] \times [0,1]$.

De två personerna möts om $|X - Y| < 1/3$ timme. Under likformig fördelning att få ett utfall i det markerade området ges som förhållandet mellan areorna. Kvadraten har area $1 \cdot 1 = 1$. Triangelnarna med kateter av längd 40 minuter = $2/3$ timme har vardera area $\frac{2}{3} \cdot \frac{2}{3} / 2 = \frac{2}{9}$. Alltså är

$$P\left(|X - Y| < \frac{1}{3}\right) = 1 - 2 \cdot \frac{2}{9} = \frac{5}{9}.$$



Paret av tider (X,Y) då de två personerna kommer till mötesplatsen är likformigt fördelat över kvadraten med sidlängd 1 timme = 60 minuter.

4.9 Den tvådimensionella stokastiska variabeln (X,Y) har simultan täthet

$$f_{X,Y}(x,y) = \frac{1}{1+c}(xy+c)e^{-(x+y)}$$

för $x \geq 0$, $y \geq 0$.

a) Marginalfördelningarna bestäms ur

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \int_0^{\infty} \frac{1}{1+c} (xy+c)e^{-(x+y)} dy = \frac{1}{1+c} e^{-x} \int_0^{\infty} (xy+c)e^{-y} dy \\ &= \frac{1}{1+c} e^{-x} \left(\underbrace{[(xy+c)e^{-y}(-1)]_0^{\infty}}_{=c} + \int_0^{\infty} xe^{-y} dy \right) = \frac{1}{1+c} e^{-x} \left(c + \underbrace{[xe^{-y}(-1)]_0^{\infty}}_{=x} \right) \\ &= \frac{c+x}{c+1} e^{-x}. \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx = \int_0^{\infty} \frac{1}{1+c} (xy+c)e^{-(x+y)} dx = \frac{c+y}{c+1} e^{-y}. \end{aligned}$$

b) Då $c = 0$ är

$$f_X(x) \cdot f_Y(y) = xe^{-x} \cdot ye^{-y} = xy \cdot e^{-(x+y)} = f_{X,Y}(x,y)$$

för $x \geq 0$ och $y \geq 0$. Alltså är $f_X(x)f_Y(y) = f_{X,Y}(x,y)$ för alla x och y så X och Y är oberoende.

Om $c \neq 0$ så är $f_X(x)f_Y(y) \neq f_{X,Y}(x,y)$, det vill säga X och Y är ej oberoende.

4.10 För $x \geq 0$ och $y \geq 0$ är den simultana täthetsfunktionen

$$f_{X,Y}(x,y) = \frac{2}{(1+x+y)^3}.$$

Nu är

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \int_0^{\infty} \frac{2}{(1+x+y)^3} dy = \left[-\frac{1}{(1+x+y)^2} \right]_{y=0}^{y=\infty} = \frac{1}{(1+x)^2}.$$

Av symmetriskäl måste X och Y ha samma fördelning så $f_Y(y) = 1/(1+y)^2$ för $y \geq 0$. Notera att $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ så de stokastiska variablerna X och Y är beroende.

b) Den simultana fördelningsfunktionen fås för $s > 0$, $t > 0$ som

$$\begin{aligned} F_{X,Y}(s,t) &= \int_{-\infty}^s \int_{-\infty}^t f_{X,Y}(x,y) dx dy = \int_0^s \int_0^t \frac{2}{(1+x+y)^3} dx dy = \int_0^s \left[-\frac{1}{(1+x+y)^2} \right]_0^t dy \\ &= \int_0^s \frac{1}{(1+y)^2} - \frac{1}{(1+t+y)^2} dy = \left[-\frac{1}{1+y} + \frac{1}{1+t+y} \right]_0^s \\ &= 1 + \frac{1}{1+t+s} - \frac{1}{1+t} - \frac{1}{1+s}. \end{aligned}$$

De marginella fördelningsfunktionerna fås som gränsvärden

$$\begin{aligned} F_X(x) &= F_{X,Y}(x, \infty) = 1 - \frac{1}{1+x} \\ F_Y(y) &= F_{X,Y}(\infty, y) = 1 - \frac{1}{1+y}. \end{aligned}$$

Ett sätt att visa att $F_{X,Y}(x,y) > F_X(x)F_Y(y)$ (vilket visar att X och Y är beroende) är följande:

$$\begin{aligned} F_{X,Y}(x,y) - F_X(x)F_Y(y) &= \left[1 - \frac{1}{1+x} - \frac{1}{1+y} + \frac{1}{1+x+y} \right] - \left(1 - \frac{1}{1+x} \right) \left(1 - \frac{1}{1+y} \right) \\ &= \frac{1}{1+x+y} - \frac{1}{(1+x)(1+y)} = \frac{1}{1+x+y} - \frac{1}{1+x+y+xy} > 0 \end{aligned}$$

eftersom $xy > 0$ då $x > 0$ och $y > 0$.

4.11 Från den simultana täthetsfunktionen

$$f_{X,Y}(x,y) = \frac{1+x+y+cxy}{c+3} e^{-(x+y)}, \quad x \geq 0, y \geq 0,$$

bestäms marginalfördelningen med hjälp av partiell integration

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \int_0^{\infty} \frac{1+x+y+cx}{c+3} e^{-(x+y)} dy = \left[\frac{1+x+y+cx}{c+3} \cdot e^{-(x+y)}(-1) \right]_0^{\infty} \\ &+ \int_0^{\infty} \frac{1+cx}{c+3} e^{-(x+y)} dy = \frac{1+x}{c+3} e^{-x} + \left[\frac{1+cx}{c+3} e^{-(x+y)}(-1) \right]_0^{\infty} = \frac{1+x}{c+3} e^{-x} + \frac{1+cx}{c+3} e^{-x} \\ &= \frac{2+(c+1)x}{c+3} e^{-x}. \end{aligned}$$

Av symmetriskäl har X och Y samma fördelning så

$$f_Y(y) = \frac{2+(c+1)y}{c+3} e^{-y}$$

och de stokastiska variablerna är oberoende om $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Högerledet kan skrivas

$$\frac{2+(c+1)x}{c+3} e^{-x} \cdot \frac{2+(c+1)y}{c+3} e^{-y} = \frac{4+2(c+1)x+2(c+1)y+(c+1)^2xy}{(c+3)^2} e^{-(x+y)}.$$

Identifiering av koefficienter ger att för oberoende så skall

$$\begin{aligned} \frac{4}{(c+3)^2} &= \frac{1}{c+3} \\ \frac{2(c+1)}{(c+3)^2} &= \frac{1}{c+3} \\ \frac{(c+1)^2}{(c+3)^2} &= \frac{c}{c+3} \end{aligned}$$

vilket har lösningen $c = 1$.

- 4.12** Låt X beskriva tiden tills person A får napp och Y tiden tills B får napp. Modell: X och Y är oberoende och exponentialfördelade med parametrar a resp. b . Då X har $f_X(x) = ae^{-ax}$ för $x \geq 0$ och Y har $f_Y(y) = be^{-by}$ för $y \geq 0$ och X och Y är oberoende så är

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = ae^{-ax}be^{-by}$$

för $(x,y) \in [0,\infty) \times [0,\infty)$. A vinner om A får napp först, det vill säga om $X < Y$. Så $P(\text{A vinner}) = P(X < Y)$ och enligt definitionen

$$\begin{aligned} P(X < Y) &= \iint_{x < y} f_{X,Y}(x,y) dy dx = \int_0^{\infty} \int_x^{\infty} f_{X,Y}(x,y) dy dx = \int_0^{\infty} \int_x^{\infty} ae^{-ax}be^{-by} dy dx \\ &= \int_0^{\infty} ae^{-ax}b \left[e^{-by} \frac{-1}{b} \right]_x^{\infty} dx = \int_0^{\infty} ae^{-ax}e^{-bx} dx = a \left[e^{-(a+b)x} \frac{-1}{a+b} \right]_0^{\infty} = \frac{a}{a+b}. \end{aligned}$$

Om $a = 2b$ är sannolikheten $2b/(2b+b) = 2/3$.

- 4.13** Iakttagelsen att bollen nuddar nätet om bollens centrum är närmre än avståndet r från en tråd, ger att sannolikheten för att inte nudda nät kan skrivas som en kvot mellan träffyta på avstånd r från tråd och hela träffyten.

- a) Den totala träffyten har arean $a \cdot a = a^2$, medan ytan på avstånd r från tråden är $(a-2r)^2$. Alltså,

$$P(\{\text{ej nudda nät}\}) = \frac{(a-2r)^2}{a^2} = \left(1 - 2\frac{r}{a}\right)^2.$$

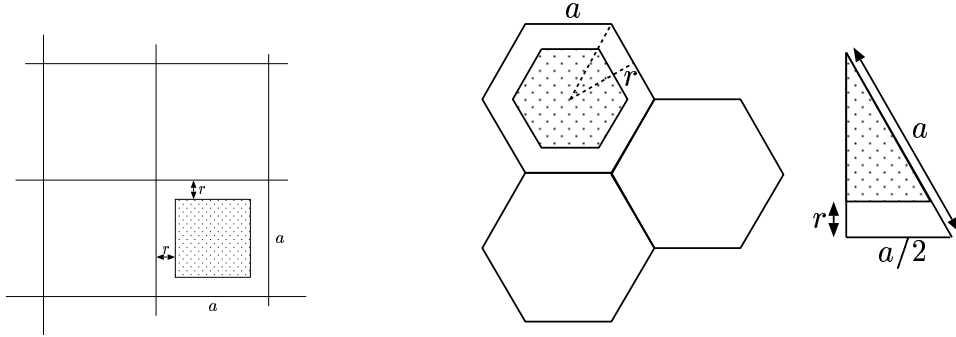
- b) En hexagon med sidlängd a kan delas in i 12 trianglar enligt figuren. Triangelns höjd är (m.h.a. Pythagoras) $\sqrt{3}a/2$ så träffyten blir $\sqrt{3}a^2/8$. Ytan på avstånd större än r från triangelns bas är

$$(\text{bas}) \cdot (\text{höjd}) \cdot \frac{1}{2} = (a/2 - \frac{r}{\sqrt{3}})(\sqrt{3}a/2 - r) \frac{1}{2}$$

så sannolikheten blir

$$P(\{\text{ej nudda nät}\}) = \frac{(a/2 - \frac{r}{\sqrt{3}})(\sqrt{3}a/2 - r)\frac{1}{2}}{\sqrt{3}a^2/8} = \left(1 - \frac{2r}{\sqrt{3}a}\right)^2.$$

(Om man inser att areorna är proportionerliga mot kvadraterna på hypotenusorna får man direkt kvoten $(a - 2r/\sqrt{3})^2/a^2$.)



4.14 Den tvådimensionella stokastiska variabeln (X, Y) har möjliga värden $\Omega_{X,Y} = \{(x, y) : x = 1, 2, \dots, y = 1, 2, \dots, x \neq y\}$.

Händelsen $(X, Y) = (x, y)$ där $y < x$ betyder att de $y - 1$ första kasten gav 3:or eller mer, vid kast y erhöles en 2:a, under kast $y + 1, \dots, x - 1$ erhöles 2:or eller mer och under kast x en 1:a. På grund av oberoende kast har sekvensen sannolikhet

$$\underbrace{\frac{4}{6} \dots \frac{4}{6}}_{y-1 \text{ st}} \cdot \frac{1}{6} \cdot \underbrace{\frac{5}{6} \dots \frac{5}{6}}_{x-1-y \text{ st}} \cdot \frac{1}{6} = \frac{4^{y-1} 5^{x-1-y}}{6^x}.$$

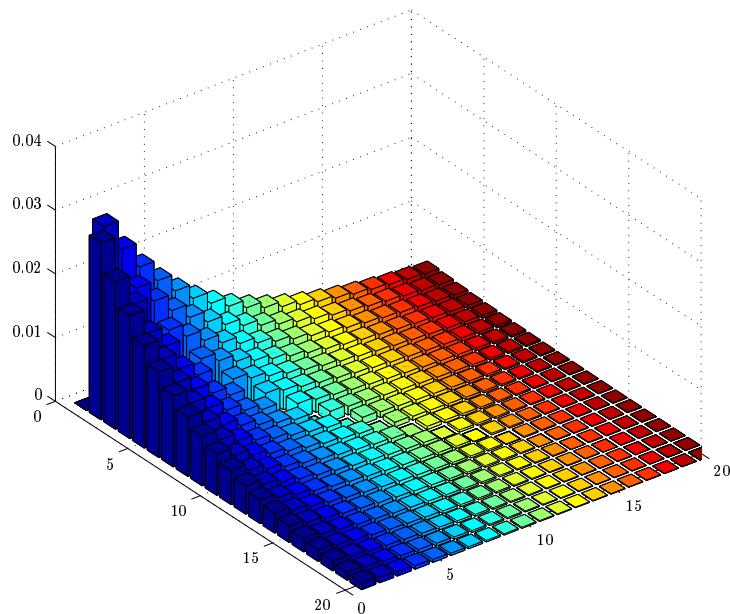
Om $x < y$ har motsvarande sekvens sannolikhet

$$\underbrace{\frac{4}{6} \dots \frac{4}{6}}_{x-1 \text{ st}} \cdot \frac{1}{6} \cdot \underbrace{\frac{5}{6} \dots \frac{5}{6}}_{y-1-x \text{ st}} \cdot \frac{1}{6} = \frac{4^{x-1} 5^{y-1-x}}{6^y}.$$

Alltså kan den simultana sannolikhetsfunktionen skrivas

$$p_{X,Y}(x, y) = \frac{4^{\min(x,y)-1} \cdot 5^{\max(x,y)-1-\min(x,y)}}{6^{\max(x,y)}} = \frac{1}{20} \left(\frac{4}{5}\right)^{\min(x,y)} \cdot \left(\frac{5}{6}\right)^{\max(x,y)}$$

för positiva heltal x och y sådana att $x \neq y$.



Den simultana fördelningen för (X, Y) , för tydlighets skull ritad som lådor istället för stolpar.

4.15 Med $Z_+ = \max(X, Y)$ där X och Y är oberoende så är

$$F_{Z_+}(t) = P(Z_+ \leq t) = P(\max(X, Y) \leq t) = P(X \leq t, Y \leq t) = P(X \leq t)P(Y \leq t) = F_X(t)F_Y(t).$$

Om X och Y har samma fördelning så är

$$F_{Z_+}(t) = F_X(t)F_Y(t) = F_X(t)^2.$$

Här är X (och Y) likformigt fördelade på intervallet $[0, a]$ så för $0 \leq t \leq a$ är

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \int_0^t \frac{1}{a} dx = \frac{t}{a}$$

och $F_{Z_+}(t) = \frac{t^2}{a^2}$.

På liknande sätt får man att med $Z_- = \min(X, Y)$ där X och Y är oberoende så är

$$\begin{aligned} F_{Z_-}(t) &= P(Z_- \leq t) = 1 - P(Z_- > t) = 1 - P(\min(X, Y) > t) = 1 - P(X > t, Y > t) \\ &= 1 - P(X > t)P(Y > t) = 1 - (1 - P(X \leq t))(1 - P(Y \leq t)) \\ &= 1 - (1 - F_X(t))(1 - F_Y(t)). \end{aligned}$$

Om X och Y har samma fördelning så är

$$F_{Z_-}(t) = 1 - (1 - F_X(t))(1 - F_Y(t)) = 1 - (1 - F_X(t))^2.$$

Här erhålls för $0 \leq t \leq a$ att

$$F_{Z_-}(t) = 1 - \left(1 - \frac{t}{a}\right)^2 = 1 - \frac{(a-t)^2}{a^2} = \frac{(2a-t)t}{a^2}.$$

4.16 De oberoende stokastiska variablerna X_1, \dots, X_n har täthetsfunktionerna

$$f_{X_i}(x) = \lambda_i e^{-\lambda_i x}, \quad x \geq 0$$

med parametrar $\lambda_1 > 0, \dots, \lambda_n > 0$.

Låt $Z = \min(X_1, \dots, X_n)$. Lösningen till problemet bygger på iakttagelsen att om $\min(X_1, \dots, X_n) > t$ så är *alla* X_1, \dots, X_n större än t . Alltså,

$$\begin{aligned} F_Z(t) &= P(Z \leq t) = 1 - P(Z > t) = 1 - P(\min(X_1, \dots, X_n) > t) \\ &= 1 - P(X_1 > t, \dots, X_n > t) = \{\text{oberoende}\} = 1 - P(X_1 > t) \cdots P(X_n > t) \\ &= 1 - (1 - P(X_1 \leq t)) \cdots (1 - P(X_n \leq t)) = 1 - (1 - F_{X_1}(t)) \cdots (1 - F_{X_n}(t)). \end{aligned}$$

För en exponentialfördelad stokastisk variabel är fördelningsfunktionen för $t \geq 0$

$$F_{X_i}(t) = \int_{-\infty}^t f_{X_i}(t) dx = \int_0^t \lambda_i e^{-\lambda_i x} dx = [(-1)e^{-\lambda_i x}]_0^t = 1 - e^{-\lambda_i t}.$$

Alltså är

$$F_Z(t) = 1 - (1 - F_{X_1}(t)) \cdots (1 - F_{X_n}(t)) = 1 - e^{-\lambda_1 t} \cdots e^{-\lambda_n t} = 1 - e^{-\lambda t}$$

där $\lambda = \sum_{i=1}^n \lambda_i$, den totala intensiteten. Sålunda är

$$f_Z(t) = \frac{d}{dt} F_Z(t) = \lambda e^{-\lambda t}, \quad t \geq 0,$$

det vill säga $Z = \min(X_1, \dots, X_n)$ är exponentialfördelad.

4.17 Med X_1, \dots, X_n som de oberoende exponentialfördelade stokastiska variabler som beskriver lampornas brinntider ges belysningens brinntid av $Y = \min(X_1, \dots, X_n)$. Samma lösning som i uppgift 4.16 ger att Y är exponentialfördelad med intensitet $\lambda_Y = \sum_{i=1}^n \lambda_{X_i} = \sum_{i=1}^n \lambda = n\lambda$, det vill säga

$$f_Z(t) = \lambda_Y e^{-\lambda_Y t} = n\lambda e^{-n\lambda t} = 20\lambda e^{-20\lambda t}, \quad t \geq 0.$$

4.18 En stokastisk variabel X med täthetsfunktion

$$f_X(x) = \frac{125 - x}{450}, \quad 95 \leq x \leq 125,$$

har fördelningsfunktion för $t \in [95, 125]$

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx = \int_{95}^t \frac{125 - x}{450} dx = \left[\frac{250x - x^2}{900} \right]_{95}^t = \frac{250t - t^2 - 14725}{900}.$$

Med $n = 8$ löpare, vars tider X_1, \dots, X_n beskrivs av oberoende stokastiska variabler, är tiden då den sista kommer i mål

$$Z = \max(X_1, \dots, X_n)$$

och tiden då den första kommer i mål $Y = \min(X_1, \dots, X_n)$.

Med Z enligt ovan så har Z fördelningsfunktion

$$\begin{aligned} F_Z(t) &= P(Z \leq t) = P(\max(X_1, \dots, X_n) \leq t) = P(X_1 \leq t, \dots, X_n \leq t) = \{\text{oberoende}\} \\ &= P(X_1 \leq t) \cdots P(X_n \leq t) = F_{X_1}(t) \cdots F_{X_n}(t). \end{aligned}$$

På samma sätt har Y fördelningsfunktion

$$\begin{aligned} F_Y(t) &= P(Y \leq t) = 1 - P(Y > t) = 1 - P(\min(X_1, \dots, X_n) > t) \\ &= 1 - P(X_1 > t, \dots, X_n > t) = \{\text{oberoende}\} = 1 - P(X_1 > t) \cdots P(X_n > t) \\ &= 1 - (1 - P(X_1 \leq t)) \cdots (1 - P(X_n \leq t)) = 1 - (1 - F_{X_1}(t)) \cdots (1 - F_{X_n}(t)). \end{aligned}$$

Nu är händelsen {alla i mål efter 100 minuter} = $\{Z \leq 100\}$ så den sökta sannolikheten är

$$P(Z \leq 100) = F_Z(100) = F_{X_1}(100) \cdots F_{X_n}(100) = [F_X(100)]^8 = \left(\frac{11}{36}\right)^8 \approx 7.60 \cdot 10^{-5}.$$

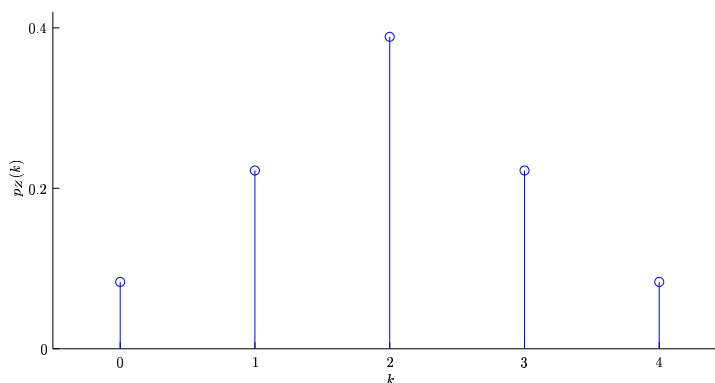
På liknande sätt är händelsen {ingen i mål efter 100 minuter} = $\{Y > 100\}$ så den sökta sannolikheten är

$$P(Y > 100) = 1 - F_Y(100) = (1 - F_{X_1}(100)) \cdots (1 - F_{X_n}(100)) = (1 - F_X(100))^8 = \left(\frac{25}{36}\right)^8 \approx 0.0541.$$

4.19 Om X och Y är stokastiska variabler med möjliga värden $\Omega_X = \{0, 1, 2\}$ respektive $\Omega_Y = \{0, 1, 2\}$ så ges de möjliga värdena på Z av $\Omega_Z = \{0, 1, 2, 3, 4\}$.

För dessa värden har vi att

$$\begin{aligned} p_Z(0) &= P(Z = 0) = P(X + Y = 0) = P(X = 0, Y = 0) = P(X = 0)P(Y = 0) = \frac{3}{36} \\ p_Z(1) &= P(Z = 1) = P(X + Y = 1) = P(\{X = 0, Y = 1\} \cup \{X = 1, Y = 0\}) \\ &= P(X = 0)P(Y = 1) + P(X = 1)P(Y = 0) = \frac{1}{18} + \frac{1}{6} = \frac{8}{36} \\ p_Z(2) &= P(Z = 2) = P(X + Y = 2) = P(\{X = 0, Y = 2\} \cup \{X = 1, Y = 1\} \cup \{X = 2, Y = 0\}) \\ &= P(X = 0)P(Y = 2) + P(X = 1)P(Y = 1) + P(X = 2)P(Y = 0) = \frac{14}{36} \\ p_Z(3) &= P(Z = 3) = P(X + Y = 3) = P(\{X = 1, Y = 2\} \cup \{X = 2, Y = 1\}) \\ &= P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1) = \frac{8}{36} \\ p_Z(4) &= P(Z = 4) = P(X + Y = 4) = P(X = 2, Y = 2) = P(X = 2)P(Y = 2) = \frac{3}{36}. \end{aligned}$$



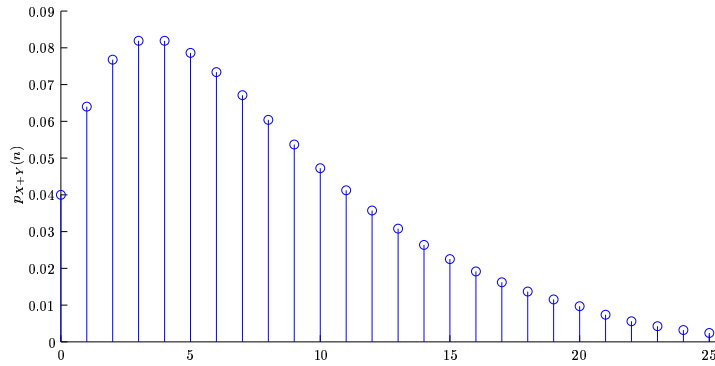
Sannolikhetsfunktionen $p_Z(k)$ för $Z = X + Y$.

4.20 Låt X och Y vara oberoende och geometriskt fördelade med parameter p , det vill säga

$$p_X(k) = p_Y(k) = (1 - p)^k p, \quad k = 0, 1, 2, \dots$$

Då ges de möjliga värdena på $Z = X + Y$ av $\Omega_Z = \{0, 1, 2, \dots\}$ där för $n \in \Omega_Z$ är

$$\begin{aligned} p_Z(n) &= P(Z = n) = P(X + Y = n) = P\left(\bigcup_k \{X = k, Y = n - k\}\right) \\ &= \sum_k P(X = k)P(Y = n - k) = \sum_{k=0}^n (1 - p)^k p \cdot (1 - p)^{n-k} p = (n + 1)(1 - p)^n p^2. \end{aligned}$$



Sannolikhetsfunktionen för $X+Y$ där X och Y är oberoende Geo(0.2)-fördelade stokastiska variabler.

4.21 Låt p vara sannolikheten att spelare A vinner en spelomgång

Om X beskriver antalet spelomgångar som spelare A vinner av $n = 5$ oberoende spelomgångar ges möjliga värden på X av $\Omega_X = \{0, 1, \dots, 5\}$. Att händelsen $\{X = k\}$ inträffar betyder att man sett en sekvens av n spelomgångar varav k vanns av spelare A, och om vi antar att i varje spelomgång vinner någon av spelarna, så vanns $n - k$ spelomgångar av spelare B. En sekvens av k A och $n - k$ B

$$\underbrace{AAA \cdots A}_{k \text{ st}} \underbrace{BB \cdots B}_{n-k \text{ st}}$$

har sannolikhet $p^k(1-p)^{n-k}$. Vi har $\binom{n}{k}$ sådana sekvenser så

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

för $k = 0, 1, \dots, n$.

Om A och B har samma vinstsannolikhet, det vill säga $p = 1 - p = 1/2$ så är

$$P(X = k) = \binom{n}{k} (1/2)^k (1/2)^{n-k} = \binom{n}{k} 2^{-n}$$

för $k = 0, 1, \dots, n$.

Med Y som antalet spelomgångar som spelare B vinner är $Y = n - X$ och $Z = X - Y = X - (n - X) = 2X - n$, det vill säga möjliga värden på $Z = X - Y$ ges av $\Omega_Z = \{-n, -n+2, \dots, n-2, n\} = \{-5, -3, -1, 1, 3, 5\}$ och för $k \in \Omega_Z$ är

$$p_Z(k) = P(Z = k) = P(2X - n = k) = P\left(X = \frac{k+n}{2}\right) = \binom{n}{\frac{k+n}{2}} 2^{-n}.$$

c) Den sökta skillnaden är $|Z|$ där de möjliga värdena är $\Omega_{|Z|} = \{1, 3, 5\}$ och för $k \in \Omega_{|Z|}$ är

$$P(|Z| = k) = P(Z = k) + P(Z = -k) = \binom{n}{\frac{k+n}{2}} 2^{-n} + \underbrace{\binom{n}{\frac{-k+n}{2}} 2^{-n}}_{=\binom{n}{(n+k)/2}} = 2 \binom{n}{\frac{k+n}{2}} 2^{-n} = \binom{n}{\frac{k+n}{2}} 2^{-4}.$$

4.22 De stokastiska variablerna X och Y är oberoende och likformigt fördelade på talen $\{1, 2, \dots, n\}$.

Då ges de möjliga värdena på $X - Y$ av $\Omega_{X-Y} = \{-n+1, -n+2, \dots, n-2, n-1\}$ och för $|X - Y|$ är motsvarande $\Omega_{|X-Y|} = \{0, 1, \dots, n-2, n-1\}$. För $k > 0$ ur denna mängd är

$$\begin{aligned} P(|X - Y| = k) &= P(X - Y = k) + P(X - Y = -k) = P(X - Y = k) + P(Y - X = k) \\ &= 2P(X - Y = k) \end{aligned}$$

eftersom X och Y har samma fördelning. För $k = 0$ så är

$$P(|X - Y| = 0) = P(X = Y) = P\left(\bigcup_i \{X = i, Y = i\}\right) = \sum_i P(X = i) P(Y = i) = \sum_{i=1}^n \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n}.$$

För $k = 1, 2, \dots, n - 1$ är

$$P(X - Y = k) = \sum_i P(X = i, Y = i - k) = \sum_i P(X = i) P(Y = i - k) = \sum_{i=1+k}^n \frac{1}{n} \cdot \frac{1}{n} = \frac{n - k}{n^2}$$

så

$$p_{|X+Y|}(k) = \begin{cases} \frac{1}{n} & \text{om } k = 0 \\ \frac{2(n-k)}{n^2} & \text{om } k = 1, 2, \dots, n - 1 \\ 0 & \text{för övrigt.} \end{cases}$$

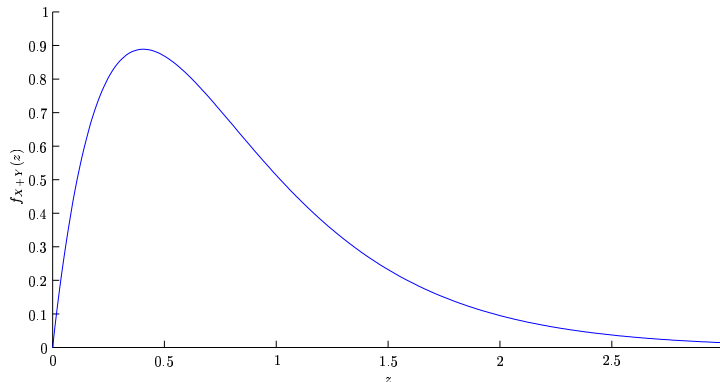
4.23 Låt X och Y vara oberoende stokastiska variabler med tätheter

$$f_X(x) = 2e^{-2x} \quad f_Y(y) = 3e^{-3y}$$

för $x \geq 0, y \geq 0$.

Summan $X + Y$ har möjliga värden $[0, \infty)$ och för $z \geq 0$ är

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_{X,Y}(u, z - u) du = \int_0^z f_X(u) f_Y(z - u) du = \int_0^z 2e^{-2u} 3e^{-3(z-u)} du \\ &= 6e^{-3z} \int_0^z e^u du = 6e^{-3z} [e^u]_0^z = 6e^{-3z} [e^z - 1]. \end{aligned}$$



Täthetsfunktionen för $X + Y$ där X och Y är oberoende och exponentialfördelade med olika intensiteter.

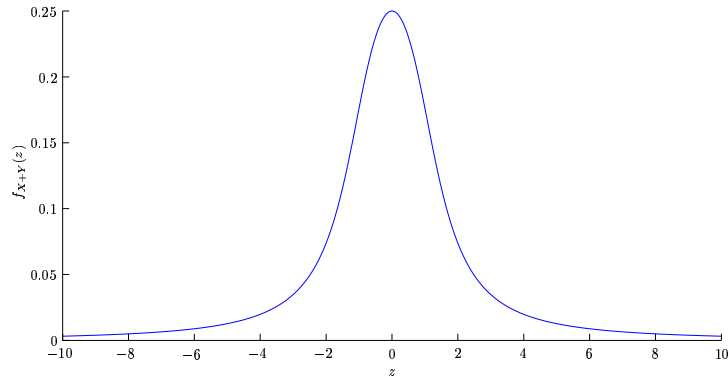
4.24 Låt X och Y vara oberoende stokastiska variabler med tätheter

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} \text{ (Cauchy)} \quad f_Y(y) = \frac{1}{2} \text{ (Likformig)}$$

för $-\infty < x < \infty$ och $-1 < y < 1$.

De möjliga värdena på $X + Y$ ges av $\Omega_{X+Y} = (-\infty, \infty)$ och

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_{X,Y}(z - u, u) du = \int_{-1}^1 f_X(z - u) f_Y(u) du = \int_{-1}^1 \frac{1}{\pi} \frac{1}{1 + (z - u)^2} \cdot \frac{1}{2} du \\ &= \frac{1}{2\pi} [-\tan^{-1}(z - u)]_{-1}^1 = \frac{1}{2\pi} [\tan^{-1}(z + 1) - \tan^{-1}(z - 1)]. \end{aligned}$$



Täthetsfunktionen för $X+Y$ där X och Y är oberoende och X är Cauchyfördelad och Y är likformigt fördelad på intervallet $(-1, 1)$.

4.25 Låt X och Y vara oberoende stokastiska variabler med tätheter

$$f_X(x) = e^{-x} \text{ (Exponential)} \quad f_Y(y) = 1 \text{ (Likformig)}$$

för $x \geq 0$ och $0 \leq y \leq 1$. De möjliga värdena på $X + Y$ ges av $\Omega_{X+Y} = [0, \infty)$ och

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-u, u) du = \int_0^{\min(1,z)} f_X(z-u) f_Y(u) du$$

Om $z > 1$ så är

$$f_{X+Y}(z) = \int_0^1 e^{-(z-u)} \cdot 1 du = e^{-z} [e^u]_0^1 = [e-1] e^{-z},$$

annars för $0 \leq z \leq 1$ är

$$f_{X+Y}(z) = \int_0^z e^{-(z-u)} \cdot 1 du = e^{-z} [e^u]_0^z = 1 - e^{-z}.$$

Dessa två uttryck kan sammantaget skrivas som

$$f_{X+Y}(z) = [e^{\min(z,1)} - 1] e^{-z}.$$

4.26 Tåget ankommer stationen X minuter efter 12:00 och lämnar stationen $X + Y$ minuter efter 12:00, dock tidigast 12:07.

Låt X och Y vara oberoende stokastiska variabler där X är likformigt fördelad på $[-2, 3]$ och Y är likformigt fördelad på $[3, 5]$, det vill säga X och Y har tätheter

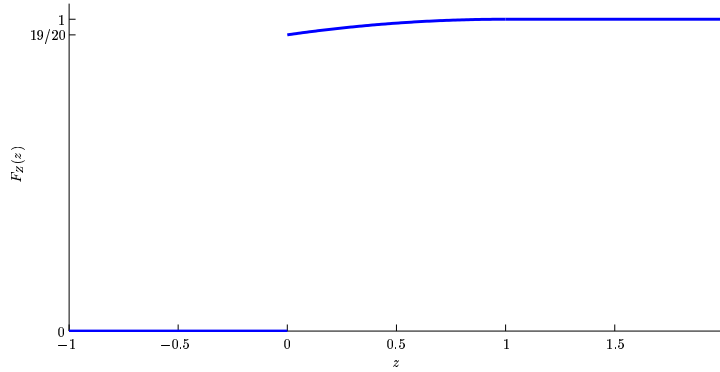
$$f_X(x) = \frac{1}{5}, \quad -2 \leq x \leq 3 \quad f_Y(y) = \frac{1}{2}, \quad 3 \leq y \leq 5. \quad \Rightarrow \quad f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{10}.$$

Om $X + Y \leq 7$ kommer tåget att avgå i tid klockan 12:07, det vill säga förseningen Z kan skrivas $Z = \max(X + Y - 7, 0)$. Möjliga värden på Z ges av intervallet $\Omega_Z = [0, 1]$. För $0 \leq z \leq 1$ så uppfyller förseningen Z

$$\begin{aligned} P(Z > z) &= P(X + Y - 7 > z) = \int \int_{(x,y): x+y > z+7} f_{X,Y}(x, y) dy dx = \int_{z+2}^3 \int_{z+7-x}^5 \frac{1}{10} dy dx \\ &= \int_{z+2}^3 \frac{x - (z+2)}{10} dx = \frac{1}{10} \left[\frac{x^2}{2} + (z+2)x \right]_{z+2}^3 = \frac{(1-z)^2}{20} \end{aligned}$$

så

$$F_Z(z) = 1 - P(Z > z) = \begin{cases} 0 & \text{om } z < 0 \\ 1 - \frac{(1-z)^2}{20} & \text{om } 0 \leq z < 1 \\ 1 & \text{om } z \geq 1. \end{cases}$$



Fördelningsfunktionen för förseningen Z . Notera att detta är en blandfördelning med en diskret komponent, fördelningsfunktionen gör ett språng i $z = 0$ motsvarande $P(Z = 0)$, och en kontinuerliga komponent där $F_Z(z)$ växer kontinuerligt.

4.27 Låt X och Y vara oberoende och Poissonfördelade med parametrar μ_A respektive μ_B . Eftersom $X \geq 0$ och $Y \geq 0$ så är de möjliga värdena på Y , betingat att $X + Y = n$, $0, 1, \dots, n$. För $k = 0, 1, \dots, n$ är enligt definitionen av betingad sannolikhet och oberoende stokastiska variabler

$$P(Y = k | X + Y = n) = \frac{P(Y = k, X + Y = n)}{P(X + Y = n)} = \frac{P(Y = k, X = n - k)}{P(X + Y = n)} = \frac{P(Y = k) P(X = n - k)}{P(X + Y = n)}.$$

Fördelningen för $X + Y$ fås med hjälp Binomialteoremet

$$\begin{aligned} P(X + Y = n) &= P\left(\bigcup_k \{X = k, Y = n - k\}\right) = \sum_k P(X = k, Y = n - k) = \{\text{oberoende}\} \\ &= \sum_k P(X = k) P(Y = n - k) = \sum_{k=0}^n \frac{\mu_B^k}{k!} e^{-\mu_B} \cdot \frac{\mu_A^{n-k}}{(n-k)!} e^{-\mu_A} \\ &= \frac{1}{n!} e^{-(\mu_A + \mu_B)} \underbrace{\sum_{k=0}^n \frac{n!}{k!(n-k)!} \mu_B^k \mu_A^{n-k}}_{=(\mu_A + \mu_B)^n} = \frac{(\mu_A + \mu_B)^n}{n!} e^{-(\mu_A + \mu_B)} = \frac{\mu^n}{n!} e^{-\mu}, \end{aligned}$$

det vill säga $X + Y$ är Poissonfördelad med parameter $\mu = \mu_A + \mu_B$. Vidare så är enligt definitionen av betingad sannolikhet

$$\begin{aligned} P(Y = k | X + Y = n) &= \frac{\frac{\mu_B^k}{k!} e^{-\mu_B} \frac{\mu_A^{n-k}}{(n-k)!} e^{-\mu_A}}{\frac{(\mu_A + \mu_B)^n}{n!} e^{-(\mu_A + \mu_B)}} = \frac{n!}{k!(n-k)!} \frac{\mu_B^k \mu_A^{n-k}}{(\mu_A + \mu_B)^n} \\ &= \binom{n}{k} \left(\frac{\mu_B}{\mu_A + \mu_B}\right)^k \left(\frac{\mu_A}{\mu_A + \mu_B}\right)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

det vill säga den betingade fördelningen för Y , givet $X + Y = n$, är $\text{Bin}(n, p)$, där $p = \mu_B / (\mu_A + \mu_B)$.

4.28 En partikel registreras med sannolikhet p oberoende av tidigare partiklar. Med X som antalet utsöndrade partiklar, X är Poissonfördelad med parameter μ , så uppfyller antalet registrerade partiklar Y

$$P(Y = k | X = n) = \binom{n}{k} p^k (1-p)^{n-k},$$

det vill säga Y är, betingat $X = n$, Binomialfördelad.

Med lagen om total sannolikhet fås

$$\begin{aligned} P(Y = k) &= \sum_n P(Y = k|X = n) P(X = n) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{\mu^n}{n!} e^{-\mu} \\ &= \frac{(\mu p)^k}{k!} e^{-\mu} \sum_{n=k}^{\infty} \frac{1}{(n-k)!} (\mu(1-p))^{n-k} = \frac{(\mu p)^k}{k!} e^{-\mu} \underbrace{\sum_{n=0}^{\infty} \frac{(\mu(1-p))^n}{n!}}_{=e^{\mu(1-p)}} \\ &= \frac{(\mu p)^k}{k!} e^{-\mu} e^{\mu(1-p)} = \frac{(\mu p)^k}{k!} e^{-\mu p}, \end{aligned}$$

det vill säga Y är Poissonfördelad med parameter μp .

4.29 Marginalfördelningarna bestäms för $x > 0$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \int_x^{\infty} e^{-y} dy = [-e^{-y}]_x^{\infty} = e^{-x},$$

det vill säga exponentialfördelad med intensitet 1, och

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx = \int_0^y e^{-y} dx = ye^{-y}$$

(en Gammafördelning).

Den betingade tätheten för Y , givet $X = x$, ges för $0 \leq x \leq y$ av

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}.$$

c) Den betingade tätheten för $Y - x$, givet $X = x$, ges för $y \geq 0$ av

$$f_{Y-x|X=x}(y) = \frac{f_{X,Y}(x, x+y)}{f_X(x)} = \frac{e^{-(y+x)}}{e^{-x}} = e^{-y},$$

det vill säga $Y - x$ är exponentialfördelad, givet $X = x$. Sålunda,

$$f_{Y-X}(y) = \int_{-\infty}^{\infty} f_{Y-X|X=x}(y) f_X(x) dx = \int_0^{\infty} e^{-y} f_X(x) dx = e^{-y} \int_0^{\infty} f_X(x) dx = e^{-y}$$

skillnaden $Y - X$ är exponentialfördelad med parameter 1.

Låt X_1 och X_2 vara oberoende och exponentialfördelade med parameter 1. Låt $X = X_1$ och $Y = X_1 + X_2$. Då är den simultana fördelningen för $0 \leq x \leq y$

$$f_{X,Y}(x,y) = f_{X_1, X_1+X_2}(x,y) = f_{X_1, X_2}(x, y-x) = \{\text{oberoende}\} = f_{X_1}(x) f_{X_2}(y-x) = e^{-x} e^{-(y-x)} = e^{-y}.$$

4.30 Låt X vara (den stokastiska) sannolikheten att apparaten går sönder under garantitiden. Möjliga värden på X är $\Omega_X = [0, 1]$ och

$$f_X(x) = 2(1-x) \quad 0 \leq x \leq 1.$$

Låt A vara händelsen att en på måfå vald apparat håller garantitiden ut. Då är

$$\begin{aligned} P(A) &= \int_{-\infty}^{\infty} P(A|X = x) f_X(x) dx = \int_0^1 (1-x) \cdot 2(1-x) dx = 2 \int_0^1 (1-x)^2 dx = 2 \left[\frac{-(1-x)^3}{3} \right]_0^1 \\ &= \frac{2}{3}. \end{aligned}$$

5.1 Lotterna i lotteriet fördelas enligt

vinstbelopp (k kr)	0	5	20	100
antal lotter	964	30	5	1
$P(X = k)$	0.964	0.030	0.005	0.001

Låt X beskriva vinstbeloppet av en på måfå vald lott. De möjliga värdena på X är $\Omega_X = \{0, 5, 20, 100\}$. Det förväntade vinstbeloppet är

$$\begin{aligned} E(X) &= \sum_k kP(X = k) \\ &= 0 \cdot P(X = 0) + 5 \cdot P(X = 5) + 20 \cdot P(X = 20) + 100 \cdot P(X = 100) \\ &= 0.35. \end{aligned}$$

Variansen kan beräknas på två sätt:

Alternativ 1: Ur definitionen

$$\begin{aligned} V(X) &= E((X - E(X))^2) = \sum_k (k - 0.35)^2 P(X = k) \\ &= (0 - 0.35)^2 \cdot P(X = 0) + (5 - 0.35)^2 \cdot P(X = 5) + (20 - 0.35)^2 \cdot P(X = 20) \\ &\quad + (100 - 0.35)^2 \cdot P(X = 100) \\ &= 12.628. \end{aligned}$$

Alternativ 2: Via

$$\begin{aligned} E(X^2) &= \sum_k k^2 P(X = k) \\ &= 0^2 \cdot P(X = 0) + 5^2 \cdot P(X = 5) + 20^2 \cdot P(X = 20) + 100^2 \cdot P(X = 100) \\ &= 12.75 \end{aligned}$$

och sedan

$$V(X) = E(X^2) - (E(X))^2 = 12.75 - 0.35^2 = 12.628.$$

Standardavvikelsen fås till $D(X) = \sqrt{V(X)} = 3.5535$.

Nettovinsten Y ges av

$$Y = X - 0.50.$$

Väntevärdet och variansen för Y kan bestämmas på tre sätt.

Alternativ 1: Ur fördelningen för nettovinsten Y . De möjliga värdena är $\{-0.5, 4.5, 19.5, 99.5\}$ med sannolikheter 0.964, 0.030, 0.005, 0.001 respektive.

$$\begin{aligned} E(Y) &= \sum_k kP(Y = k) \\ &= (-0.5) \cdot P(Y = -0.5) + 4.5 \cdot P(Y = 4.5) + 19.5 \cdot P(Y = 19.5) \\ &\quad + 99.5 \cdot P(Y = 99.5) = -0.15 \end{aligned}$$

och

$$\begin{aligned} V(Y) &= E((Y - E(Y))^2) = \sum_k (k + 0.15)^2 P(Y = k) \\ &= (-0.5 + 0.15)^2 \cdot P(Y = -0.5) + \dots + (99.5 + 0.15)^2 \cdot P(Y = 99.5) = 12.628. \end{aligned}$$

Alternativ 2: Via fördelningen för X .

$$\begin{aligned} E(Y) &= E(X - 0.50) = \sum_k (k - 0.50)P(X = k) \\ &= (0 - 0.50)P(X = 0) + \cdots + (100 - 0.50)P(X = 100) = -0.15 \end{aligned}$$

och

$$\begin{aligned} V(Y) &= E((Y - E(Y))^2) = E(((X - 0.5) + 0.15)^2) = E((X - 0.35)^2) \\ &= \sum_k (k - 0.35)^2 P(X = k) = 12.628. \end{aligned}$$

Alternativ 3: Via lineariteten för väntevärden.

$$E(Y) = E(X - 0.50) = E(X) - 0.50 = 0.35 - 0.50 = -0.15$$

och

$$V(Y) = V(X - 0.50) = V(X) = 12.628.$$

5.2 Uppgiften kan lösas på två sätt.

Alternativ 1: Låt X beskriva antalet steg som en löpare flyttar. Då är

tärningsresultat	1	2	3	4	5	6
antal steg	6	2	3	4	5	6

De möjliga värdena på X ges av $\Omega_X = \{2, 3, 4, 5, 6\}$ där $p_X(k) = P(X = k) = 1/6$ för $k = 2, \dots, 5$ och $p_X(6) = P(X = 6) = 1/6 + 1/6 = 1/3$.

Väntevärdet för X är enligt definitionen

$$E(X) = \sum_k k p_X(k) = 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \cdots + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{3} = \frac{13}{3}.$$

Alternativ 2: Låt X beskriva resultatet av tärningskastet. Möjliga värden på X är $\{1, 2, \dots, 6\}$ och $p_X(k) = 1/6$ för $k = 1, \dots, 6$. Låt $g(x)$ vara funktionen $g: \Omega_X \rightarrow \{2, 3, \dots, 6\}$ given av:

tärningsresultat (x)	1	2	3	4	5	6
antal steg $g(x)$	6	2	3	4	5	6

Då är

$$E(g(X)) = \sum_k g(k) p_X(k) = 6 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{13}{3}.$$

5.3 Mätinstrumentet ger ett mätfel X med täthet $f_X(x) = 100(1 - 100|x|)$ för $-0.01 \leq x \leq 0.01$. Då är

$$\begin{aligned} E(X) &= \int x f_X(x) dx = \int_{-0.01}^{0.01} x \cdot 100(1 - 100|x|) dx \\ &= \{\text{Symmetriskt intervall, udda integrand}\} = 0. \end{aligned}$$

Vidare så är

$$\begin{aligned} V(X) &= E((X - E(X))^2) = E(X^2) = \int x^2 f_X(x) dx = \int_{-0.01}^{0.01} x^2 \cdot 100(1 - 100|x|) dx \\ &= \{\text{Symmetriskt intervall, jämn integrand}\} = 2 \int_0^{0.01} x^2 \cdot 100(1 - 100x) dx \\ &= 200 \left[\frac{x^3}{3} - 100 \frac{x^4}{4} \right]_0^{0.01} = \frac{10^{-4}}{6} \end{aligned}$$

så

$$D(X) = \sqrt{V(X)} = 10^{-2}/\sqrt{6}.$$

5.4 Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion $f_X(x) = 2x/a^2$, $0 \leq x \leq a$. Då är

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^a x \cdot \frac{2x}{a^2} = \frac{2}{a^2} \left[\frac{x^3}{3} \right]_0^a = \frac{2}{a^2} \cdot \left[\frac{a^3 - 0}{3} \right] = \frac{2}{3} a.$$

5.5 Låt X vara en diskret stokastisk variabel med möjliga värden givna av $\Omega_X = \{1, 2, 3, \dots\}$ och

$$p_X(k) = P(X = k) = \frac{6}{\pi^2} \cdot \frac{1}{k^2}, \quad k = 1, 2, \dots$$

Notera att detta är en giltig fördelning: $p_X(k) \geq 0$ för $k \in \Omega_X$ och

$$\sum_{k \in \Omega_X} p_X(k) = \sum_{k=1}^{\infty} \frac{6}{\pi^2} \cdot \frac{1}{k^2} = \frac{6}{\pi^2} \underbrace{\sum_{k=1}^{\infty} \frac{1}{k^2}}_{=\pi^2/6} = \frac{6}{\pi^2} \cdot \frac{\pi^2}{6} = 1.$$

För denna fördelning är

$$\sum_{k \in \Omega_X} k p_X(k) = \sum_{k=1}^{\infty} k \cdot \frac{6}{\pi^2 k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$

eftersom den harmoniska serien är divergent, alltså existerar inte väntevärdet $E(X)$ som annars hade definierats av vänsterledet.

5.6 Låt X vara tiden som en bil parkeras i parkeringshuset. De möjliga värdena för X ges av intervallet $\Omega_X = [0, \infty)$ och X har täthet $f_X(x) = e^{-x}$, det vill säga X är exponentialfördelad med parameter 1.

Låt Y beskriva kostnaden för att parkera en bil. Då är $Y = 10 + 5X$. Här följer tre alternativ för att bestämma $E(Y)$.

Alternativ 1: Via fördelningen för Y . De möjliga värdena för $Y = 10 + 5X$ ges av $\Omega_Y = [10, \infty)$. För $t \in \Omega_Y$ så är fördelningsfunktionen för Y

$$F_Y(t) = P(Y \leq t) = P(10 + 5X \leq t) = P(X \leq (t - 10)/5) = F_X((t - 10)/5).$$

Deriveras denna fås tätheten

$$f_Y(t) = \frac{d}{dt} F_Y(t) = \frac{d}{dt} F_X((t - 10)/5) = f_X((t - 10)/5) \frac{1}{5} = \frac{1}{5} e^{-\frac{1}{5}(t-10)}.$$

för $t \geq 10$. Med denna täthet bestäms väntevärdet medelst partiell integration.

$$\begin{aligned} E(Y) &= \int t f_Y(t) dt = \int_{10}^{\infty} t \frac{1}{5} e^{-\frac{1}{5}(t-10)} dt = \underbrace{\left[-t e^{-\frac{1}{5}(t-10)} \right]_{10}^{\infty}}_{=10} + \int_{10}^{\infty} e^{-\frac{1}{5}(t-10)} dt \\ &= 10 + \left[-5 e^{-\frac{1}{5}(t-10)} \right]_{10}^{\infty} = 10 + [5 - 0] = 15 \text{ kr.} \end{aligned}$$

Alternativ 2: Via fördelningen för X . Eftersom $E(g(X)) = \int g(x) f_X(x) dx$ så är via partiell integration

$$\begin{aligned} E(Y) &= E(5 + 10X) = \int_x (5 + 10x) f_X(x) dx = \int_0^{\infty} (5 + 10x) e^{-x} dx \\ &= \underbrace{\left[-(5 + 10x) e^{-x} \right]_0^{\infty}}_{=5} + \int_0^{\infty} 10 e^{-x} dx = 5 + 10 \left[-e^{-x} \right]_0^{\infty} = 5 + 10 = 15 \text{ kr.} \end{aligned}$$

Alternativ 3: Via $E(X)$ och räknelagarna för väntevärden. För X gäller att

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \cdot e^{-x} dx = \underbrace{[-xe^{-x}]_0^{\infty}}_{=0} + \int_0^{\infty} e^{-x} dx = [-e^{-x}]_0^{\infty} = 1.$$

Alltså

$$E(Y) = E(10 + 5X) = 10 + 5E(X) = 10 + 5 = 15 \text{ kr.}$$

5.7 Låt X vara en stokastisk variabel med täthet $f_X(x) = 2e^{-2x}$ för $x \geq 0$. Då är

$$E(e^X) = \int_{-\infty}^{\infty} e^x f_X(x) dx = \int_0^{\infty} e^x 2e^{-2x} dx = 2 \int_0^{\infty} e^{-x} dx = 2[-e^{-x}]_0^{\infty} = 2.$$

Kommentar: Funktionen $m_X(s) = E(e^{sX})$ kallas för den momentgenererande funktionen till den stokastiska variabeln X . Här beräknades $m_X(1)$.

5.8 Den stokastiska variabeln X har fördelningsfunktion $F_X(x) = 1 - (1+x)^{-a}$ för $x \geq 0$ och en parameter $a > 0$. Här följer två alternativ för att beräkna $E\left(\frac{1}{1+X}\right)$.

Alternativ 1: Låt $Y = \frac{1}{1+X}$. Då ges de möjliga värdena för Y av $\Omega_Y = (0, 1]$ och för $t \in \Omega_Y$ är

$$\begin{aligned} F_Y(t) &= P(Y \leq t) = P\left(\frac{1}{1+X} \leq t\right) = P\left(X \geq \frac{1}{t} - 1\right) = 1 - F_X\left(\frac{1}{t} - 1\right) \\ &= 1 - \left[1 - \left(1 + \frac{1}{t} - 1\right)^{-a}\right] = t^a. \end{aligned}$$

Då har Y täthetsfunktion

$$f_Y(t) = \frac{d}{dt} F_Y(t) = at^{a-1}, \quad 0 < t \leq 1,$$

och väntevärde

$$E(Y) = \int_{-\infty}^{\infty} t f_Y(t) dt = \int_0^1 t \cdot at^{a-1} dt = a \left[\frac{t^{a+1}}{a+1} \right]_0^1 = \frac{a}{a+1}.$$

Alternativ 2: Deriveras fördelningsfunktionen för X fås

$$f_X(x) = \frac{d}{dx} F_X(x) = 0 - (-a)(1+x)^{-a-1} = a(1+x)^{-a-1}, \quad x \geq 0.$$

Nu kan väntevärdet beräknas enligt

$$\begin{aligned} E\left(\frac{1}{1+X}\right) &= \int_{-\infty}^{\infty} \frac{1}{1+x} \cdot f_X(x) dx = \int_0^{\infty} \frac{1}{1+x} \cdot \frac{a}{(1+x)^{a+1}} dx = a \int_0^{\infty} \frac{1}{(1+x)^{a+2}} dx \\ &= a \left[\frac{1}{(1+x)^{a+1}} \frac{-1}{a+1} \right]_0^{\infty} = \frac{a}{a+1}. \end{aligned}$$

5.9 Om X är en stokastisk variabel med täthetsfunktion $f_X(x) = \frac{1}{10}$ då $-5 \leq x \leq 5$, det vill säga X är likformigt fördelad på intervallet $[-5, 5]$, så är

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-5}^5 g(x) \frac{1}{10} dx = \int_{-5}^0 (-1) \frac{1}{10} dx + \int_0^5 2 \frac{1}{10} dx = 5 \cdot \frac{-1}{10} + 5 \cdot \frac{2}{10} = \frac{1}{2}.$$

5.10 Låt X beskriva mängden vätska som tappas upp. Då är

$$f_X(x) = \frac{1}{(1+x)^2}, \quad x \geq 0.$$

Mängden vätska i kärlet beskrivs av den stokastiska variabeln $g(X)$ där funktionen $g(x)$ är

$$g(x) = \begin{cases} x & \text{om } 0 \leq x \leq a \\ a & \text{om } x > a. \end{cases}$$

Alternativ 1: Den genomsnittliga vätskemängden i kärlet är då

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_0^{\infty} g(x) \frac{1}{(1+x)^2} dx \\ &= \int_0^a x \cdot \frac{1}{(1+x)^2} dx + \int_a^{\infty} a \cdot \frac{1}{(1+x)^2} dx = \int_1^{a+1} \frac{1}{y} - \frac{1}{y^2} dy + a \left[\frac{-1}{1+x} \right]_a^{\infty} \\ &= \left[\ln(y) + \frac{1}{y} \right]_1^{a+1} + \frac{a}{1+a} = \ln(a+1). \end{aligned}$$

Alternativ 2: Med $Y = g(X)$ så ges de möjliga värdena på Y av $\Omega_Y = [0, a]$. Variabeln Y är en blandfördelning med en diskret komponent,

$$P(Y = a) = P(X \geq a) = \int_a^{\infty} f_X(x) dx = \int_a^{\infty} \frac{1}{(1+x)^2} dx = \left[\frac{-1}{1+x} \right]_a^{\infty} = \frac{1}{1+a},$$

och en kontinuerlig komponent

$$f_Y(x) = f_X(x) = \frac{1}{(1+x)^2}, \quad 0 \leq x < a.$$

Väntevärdet beräknas som

$$\begin{aligned} E(Y) &= aP(Y = a) + \int_0^a x f_Y(x) dx = a \cdot \frac{1}{1+a} + \int_0^a \frac{x}{(1+x)^2} dx = \frac{a}{1+a} + \int_1^{a+1} \frac{z-1}{z^2} dz \\ &= \frac{a}{1+a} + \left[\ln(z) + \frac{1}{z} \right]_1^{a+1} = \ln(a+1). \end{aligned}$$

5.11 Om $E(X) = 81$ och $V(X) = 81$ är standardavvikelsen $D(X) = \sqrt{V(X)} = \sqrt{81} = 9$ och variationskoefficienten

$$\frac{D(X)}{E(X)} = \frac{9}{81} = \frac{1}{9} \approx 11.1\%.$$

5.12 Den stokastiska variabeln X har täthetsfunktion $f_X(x) = 3x^{-4}$ för $x \geq 1$. Då är väntevärdet

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_1^{\infty} x \cdot \frac{3}{x^4} dx = 3 \left[\frac{1}{x^2} \cdot \frac{-1}{2} \right]_1^{\infty} = \frac{3}{2}.$$

Variansen kan beräknas på två sätt.

Alternativ 1: Ur definitionen:

$$\begin{aligned} V(X) &= E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx = \int_1^{\infty} \left(x - \frac{3}{2}\right)^2 \cdot \frac{3}{x^4} dx \\ &= \int_1^{\infty} \frac{3(x^2 - 3x + \frac{9}{4})}{x^4} dx = 3 \left[\frac{-1^2}{x} - \frac{-3}{2x^2} + \frac{-3}{4x^3} \right]_1^{\infty} = \frac{3}{4}. \end{aligned}$$

Alternativ 2: Genom att först beräkna

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_1^{\infty} x^2 \cdot \frac{3}{x^4} dx = 3 \left[\frac{-1}{x} \right]_1^{\infty} = 3$$

och sedan utnyttja

$$V(X) = E(X^2) - (E(X))^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$$

5.13 Låt X ha täthetsfunktion $f_X(x) = 2x$ för $0 \leq x \leq 1$. Då är

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 2x dx = 2 \left[\frac{x^3}{3} \right]_0^1 = \frac{2}{3}.$$

Vidare,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \cdot 2x dx = 2 \left[\frac{x^4}{4} \right]_0^1 = \frac{1}{2}$$

så

$$V(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

varför $\sigma = D(X) = \sqrt{V(X)} = 1/\sqrt{18}$.

Sannolikheterna i b) och c) kan beräknas på två sätt:

Alternativ 1:

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + \sigma) &= \int_{\mu - 2\sigma}^{\mu + \sigma} f_X(x) dx = \int_{\mu - 2\sigma}^{\mu + \sigma} 2x dx = [x^2]_{\mu - 2\sigma}^{\mu + \sigma} = (\mu + \sigma)^2 - (\mu - 2\sigma)^2 \\ &= ((\mu + \sigma) - (\mu - 2\sigma))((\mu + \sigma) + (\mu - 2\sigma)) = 3\sigma(2\mu - \sigma) \\ &= \frac{3}{\sqrt{18}} \left(2 \cdot \frac{2}{3} - \frac{1}{\sqrt{18}} \right) = \frac{4\sqrt{2} - 1}{6}. \end{aligned}$$

Den andra sannolikheten bestäms på samma sätt. Notera att $\mu + 2\sigma > 1$.

$$\begin{aligned} P(\mu - \sigma < X < \mu + 2\sigma) &= \int_{\mu - \sigma}^{\mu + 2\sigma} f_X(x) dx = \int_{\mu - \sigma}^1 2x dx = [x^2]_{\mu - \sigma}^1 \\ &= 1 - (\mu - \sigma)^2 = 1 - \left(\frac{2}{3} - \frac{1}{\sqrt{18}}\right)^2 = \frac{1}{2} + \frac{2\sqrt{2}}{9}. \end{aligned}$$

Alternativ 2: Genom att först bestämma fördelningsfunktionen för X . För $t \in [0, 1]$ är

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx = \int_0^t 2x dx = [x^2]_0^t = t^2.$$

Sedan bestäms sannolikheterna enligt, notera att $0 \leq \mu - 2\sigma, \mu + \sigma \leq 1$,

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + \sigma) &= F_X(\mu + \sigma) - F_X(\mu - 2\sigma) = (\mu + \sigma)^2 - (\mu - 2\sigma)^2 \\ &= ((\mu + \sigma) - (\mu - 2\sigma))((\mu + \sigma) + (\mu - 2\sigma)) = 3\sigma(2\mu - \sigma) \\ &= \frac{3}{\sqrt{18}} \left(2 \cdot \frac{2}{3} - \frac{1}{\sqrt{18}} \right) = \frac{4\sqrt{2} - 1}{6}, \end{aligned}$$

och, notera att $0 \leq \mu - \sigma \leq 1 \leq \mu + 2\sigma$,

$$P(\mu - \sigma < X < \mu + 2\sigma) = F_X(\mu + 2\sigma) - F_X(\mu - \sigma) = 1 - (\mu - \sigma)^2 = 1 - \left(\frac{2}{3} - \frac{1}{\sqrt{18}}\right)^2 = \frac{1}{2} + \frac{2\sqrt{2}}{9}.$$

5.14 Vi beräknar $V(X^2)$ genom att utnyttja

$$V(X^2) = E((X^2)^2) - (E(X^2))^2.$$

Om X är likformigt fördelad på intervallet $[0, 1]$ så är $f_X(x) = 1$, $0 \leq x \leq 1$, och

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \cdot 1 dx = \left[\frac{x^3}{3} \right]_0^1 = \frac{1}{3}.$$

På samma sätt är

$$E(X^4) = \int_{-\infty}^{\infty} x^4 f_X(x) dx = \int_0^1 x^4 \cdot 1 dx = \left[\frac{x^5}{5} \right]_0^1 = \frac{1}{5}.$$

(Generellt är $E(X^k) = 1/(k+1)$.) Alltså är

$$V(X^2) = E(X^4) - (E(X^2))^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}.$$

5.15 Låt X beskriva resultatet vid mätningen av surhetsgrad med en pH-meter. Då är

$$X = (\text{surhetsgrad}) + (\text{mätfel}) = 5.8 + Y$$

där mätfelet Y är en stokastisk variabel med $E(Y) = d$ (det systematiska felet) och $D(Y) = \sigma$. Notera att Y kan skrivas $Y = d + \epsilon$ där det slumpmässiga felet ϵ är en stokastisk variabel med $E(\epsilon) = 0$, $D(\epsilon) = \sigma$.

Med räknelagarna för väntevärden och varianser erhålls

$$E(X) = E(5.8 + Y) = 5.8 + E(Y) = 5.8 + d = 6.2$$

om $d = 0.4$, och

$$V(X) = V(5.8 + Y) = V(Y) = \sigma^2$$

så $D(X) = \sigma = 0.05$.

5.16 Låt (X, Y) vara en tvådimensionell stokastisk variabel med simultan fördelning:

		j			
		0	1	2	$p_Y(k)$
k	1	0.1	0.2	0.3	0.6
	2	0.4	0	0	0.4
		$p_X(j)$	0.5	0.2	0.3

där marginalfördelningarna $p_X(j)$ och $p_Y(k)$ bestäms ur

$$p_X(j) = \sum_{k:(j,k) \in \Omega_{X,Y}} p_{X,Y}(j,k) \quad p_Y(k) = \sum_{j:(j,k) \in \Omega_{X,Y}} p_{X,Y}(j,k).$$

Då är

$$E(X) = \sum_j j \cdot p_X(j) = 0 \cdot 0.5 + 1 \cdot 0.2 + 2 \cdot 0.3 = 0.8.$$

och

$$E(Y) = \sum_k k \cdot p_Y(k) = 1 \cdot 0.6 + 2 \cdot 0.4 = 1.4$$

Vidare,

$$E(XY) = \sum_{(j,k)} jk \cdot p_{X,Y}(j,k) = (0 \cdot 1) \cdot 0.1 + (0 \cdot 2) \cdot 0.4 + (1 \cdot 1) \cdot 0.2 + (1 \cdot 2) \cdot 0 + (2 \cdot 1) \cdot 0.3 + (2 \cdot 2) \cdot 0 = 0.2 + 0.6 = 0.8$$

så kovariansen fås till

$$C(X, Y) = E(XY) - E(X)E(Y) = 0.8 - 0.8 \cdot 1.4 = -0.32.$$

5.17 Låt X beskriva värdet av det första och Y värdet av det andra på måfå valda myntet. Då är $\Omega_X = \Omega_Y = \{1, 5\}$ och

$$P(X = 1) = \frac{2}{3}, \quad P(X = 5) = \frac{1}{3}.$$

Nu är

$$P(Y = 1|X = 1) = \frac{1}{2} \quad P(Y = 5|X = 1) = \frac{1}{2} \quad P(Y = 1|X = 5) = 1 \quad P(Y = 5|X = 5) = 0$$

vilket med $P(X = x, Y = y) = P(Y = y|X = x)P(X = x)$ ger den simultana fördelningen:

$$\begin{array}{c}
 p_{X,Y}(j, k) \quad \frac{j}{\quad} \\
 \begin{array}{c|cc}
 & 1 & 5 \\
 \hline
 k & 1 & \boxed{\begin{array}{cc} 1/3 & 1/3 \end{array}} \\
 & 5 & \boxed{\begin{array}{cc} 1/3 & 0 \end{array}}
 \end{array}
 \quad \begin{array}{c} p_Y(k) \\ 2/3 \\ 1/3 \end{array} \\
 \\
 p_X(j) \quad 2/3 \quad 1/3
 \end{array}$$

Notera att X och Y har samma fördelning.

Nu är

$$E(X) = \sum_j j \cdot p_X(j) = 1 \cdot \frac{2}{3} + 5 \cdot \frac{1}{3} = \frac{7}{3} = E(Y)$$

och

$$E(X^2) = \sum_j j^2 \cdot p_X(j) = 1^2 \cdot \frac{2}{3} + 5^2 \cdot \frac{1}{3} = 9$$

så

$$V(X) = E(X^2) - (E(X))^2 = 9 - \left(\frac{7}{3}\right)^2 = \frac{32}{9} = V(Y).$$

Nu är

$$E(XY) = \sum_{(j,k)} jk \cdot p_{X,Y}(j, k) = (1 \cdot 1) \cdot \frac{1}{3} + (1 \cdot 5) \cdot \frac{1}{3} + (5 \cdot 1) \cdot \frac{1}{3} + (5 \cdot 5) \cdot 0 = \frac{11}{3}$$

så kovariansen fås till

$$C(X, Y) = E(XY) - E(X)E(Y) = \frac{33}{9} - \frac{49}{9} = -\frac{16}{9}.$$

Korrelationen $\rho(X, Y)$ bestäms av

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)V(Y)}} = -\frac{16/9}{32/9} = -\frac{1}{2}.$$

5.18 De möjliga värdena på (X, Y) är $(0, 1)$, $(0, -1)$, $(-1, 0)$ och $(1, 0)$ med sannolikhet $1/4$ vardera. De möjliga värdena på X är $-1, 0, 1$ med sannolikhet $1/4, 1/4 + 1/4 = 1/2$ och $1/4$ respektive. Notera att X och Y är *beroende* stokastiska variabler. Detta kan visas formellt genom att till exempel visa att

$$P(X = 1, Y = 1) = 0 \neq \frac{1}{4} \cdot \frac{1}{4} = P(X = 1)P(Y = 1).$$

Nu är

$$E(X) = \sum_k k p_X(k) = (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0.$$

Av symmetriskäl har Y samma fördelning så $E(Y) = 0$. Nu är

$$\begin{aligned} C(X, Y) &= E((X - E(X))(Y - E(Y))) = E(XY) = \sum_{i,j} (i \cdot j) P(X = i, Y = j) \\ &= (0 \cdot 1) \frac{1}{4} + (0 \cdot (-1)) \frac{1}{4} + ((-1) \cdot 0) \frac{1}{4} + (1 \cdot 0) \frac{1}{4} = 0. \end{aligned}$$

Korrelationen $\rho(X, Y) = C(X, Y) / \sqrt{V(X)V(Y)} = 0$.

5.19 Låt X beskriva positionen efter ett hopp och Y positionen efter två hopp.

Alternativ 1: De möjliga värdena på (X, Y) ges av $(-1, -2)$, $(-1, 0)$, $(1, 0)$ och $(1, 2)$ med sannolikhet $1/4$ vardera. De möjliga värdena för X är -1 och 1 med sannolikhet $1/2$ vardera medan för Y är värdena -2 , 0 och 2 med sannolikhet $1/4$, $1/2$ och $1/4$ respektive.

Nu är

$$\begin{aligned} E(X) &= \sum_k k p_X(k) = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0 \\ V(X) &= E((X - E(X))^2) = E(X^2) = \sum_k k^2 p_X(k) = 1 \\ E(Y) &= \sum_k k p_Y(k) = -2 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 0 \\ V(Y) &= E((Y - E(Y))^2) = E(Y^2) = \sum_k k^2 p_Y(k) = 2 \\ C(X, Y) &= E((X - E(X))(Y - E(Y))) = E(XY) = \sum_{i,j} (i \cdot j) \cdot P(X = i, Y = j) \\ &= (-2)(-1) \frac{1}{4} + 0(-1) \frac{1}{4} + 0 \cdot 1 \frac{1}{4} + 2 \cdot 1 \frac{1}{4} = 1. \end{aligned}$$

Alltså är

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{1}{\sqrt{1 \cdot 2}} = \frac{1}{\sqrt{2}}.$$

Alternativ 2: Låt X_1 och X_2 vara förflyttningarna i tidsteg 1 och 2. Då är

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

och X_1 och X_2 är oberoende. Vidare,

$$E(X_i) = \sum_k k P(X_i = k) = 0$$

och

$$V(X_i) = E((X_i - E(X_i))^2) = E(X_i^2) = \sum_k k^2 P(X_i = k) = 1.$$

Nu är $X = X_1$ och $Y = X_1 + X_2$ och

$$C(X, Y) = C(X_1, X_1 + X_2) = \underbrace{C(X_1, X_1)}_{=V(X_1)} + \underbrace{C(X_1, X_2)}_{=0} = V(X_1) = 1,$$

$$V(Y) = V(X_1 + X_2) = V(X_1) + V(X_2) = 2$$

så

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{1}{\sqrt{1 \cdot 2}} = \frac{1}{\sqrt{2}}.$$

5.20 Låt (X, Y) ha simultan täthet

$$f_{X,Y}(x, y) = \frac{1}{2x}e^{-x}, \quad x > 0, \quad -x \leq y \leq x$$

Eftersom de möjliga värdena på Y beror på X är de stokastiska variablerna beroende. För att visa detta formellt kan man se att obetingat ges de möjliga värdena på Y av $\Omega_Y = (-\infty, \infty)$ och motsvarande för X är $\Omega_X = [0, \infty)$. Dessa värden har positiva tätheter, dvs $f_X(x) > 0$ och $f_Y(y) > 0$ för $x \in \Omega_X$ och $y \in \Omega_Y$, men $f_{X,Y}(x, y) = 0$ då $y > x$ så alltså är $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$ vilket visar att X och Y är beroende.

Vi noterar att

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-x}^x \frac{1}{2x}e^{-x} dy = 2x \cdot \frac{1}{2x}e^{-x} = e^{-x},$$

det vill säga X är exponentialfördelad med intensitet 1. Alltså är $E(X) = 1$. Av symmetriskäl är $E(Y) = 0$.

Slutligen

$$E(XY) = \int_0^{\infty} \int_{-x}^x xy \cdot \frac{1}{2x}e^{-x} dy dx = \frac{1}{2} \int_0^{\infty} e^{-x} \underbrace{\int_{-x}^x y dy}_{=0} dx = 0$$

så $C(X, Y) = E(XY) - E(X)E(Y) = 0 - 1 \cdot 0 = 0$. Detta medför att $\rho(X, Y) = 0$ och variablerna är okorrelerade men ej oberoende.

Notera att för att lösa uppgiften behöver inte varianserna $V(X)$ och $V(Y)$ bestämmas. Vill man räkna ut dessa är $V(X) = 1$ (eftersom X är exponentialfördelad med parameter 1 och $V(Y) = E(Y^2) - (E(Y))^2 = E(Y^2)$ där

$$\begin{aligned} E(Y^2) &= \int_0^{\infty} \int_{-x}^x y^2 \frac{1}{2x}e^{-x} dy dx = \frac{1}{3} \int_0^{\infty} x^2 e^{-x} dx = \frac{1}{3} [-x^2 e^{-x}]_0^{\infty} + \frac{2}{3} \int_0^{\infty} x e^{-x} dx \\ &= 0 + \frac{2}{3} [-x e^{-x}]_0^{\infty} + \frac{2}{3} \int_0^{\infty} e^{-x} dx = \frac{2}{3} [-e^{-x}]_0^{\infty} = \frac{2}{3}, \end{aligned}$$

så $V(Y) = 2/3$. Observera att $E(Y^2)$ beräknades genom att utnyttja den simultana fördelningen $f_{X,Y}(x, y)$ och inte med marginalfördelningen $f_Y(y)$.

5.21 Låt X_i beskriva antalet test som behövs vid grupp i . Då ges de möjliga värdena på X_i av $\{1, k+1\}$ och

$$P(X_i = 1) = P(\text{Alla lampor är hela}) = (1-p)^k \quad P(X_i = k+1) = 1 - P(X_i = 1) = 1 - (1-p)^k.$$

Alltså är

$$E(X_i) = 1 \cdot (1-p)^k + (k+1) \cdot (1 - (1-p)^k) = (k+1) - k(1-p)^k.$$

Det totala antalet test som behövs är

$$X = X_1 + X_2 + \dots + X_n$$

där X_1, \dots, X_n är oberoende stokastiska variabler. Nu är

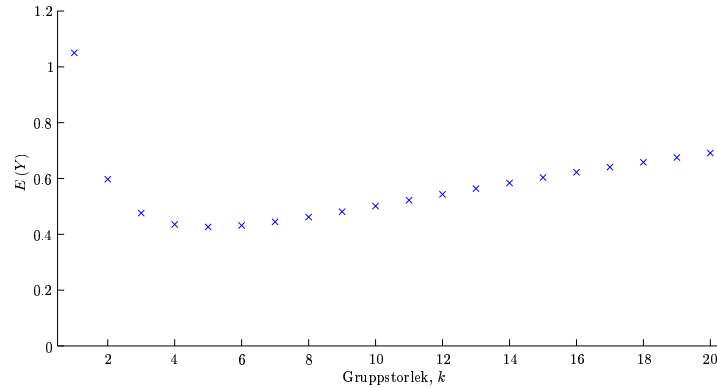
$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = n [(k+1) - k(1-p)^k]$$

och med $Y = X/(nk)$ är

$$E(Y) = E\left(\frac{1}{nk}X\right) = \frac{1}{nk}E(X) = \frac{1}{nk} \cdot n [(k+1) - k(1-p)^k] = 1 + \frac{1}{k} - (1-p)^k.$$

Bestämning av vilket k som ger minimum för givna p löstes numeriskt och följande tabell erhöles.

Felsannolikhet, p	k : $E(Y)$ minimal	$\min_k E(Y)$
0.05	5	0.42622
0.01	11	0.19557
0.001	32	0.062759
0.0001	101	0.019951



Det förväntade antalet försök som måste göras vid gruppstorlek k då $p = 0.05$. Den förväntade antalet är som lägst för $k = 5$ vilket ger $E(Y) \approx 0.426$.

5.22 Låt X_1 , X_2 och X_3 vara oberoende och $E(X_i) = 2$ och $D(X_i) = 3$ för $i = 1, 2, 3$. Med $Y = 3X_1 - 2X_2 + X_3 - 6$ så är

$$E(Y) = E(3X_1 - 2X_2 + X_3 - 6) = 3E(X_1) - 2E(X_2) + E(X_3) - 6 = 3 \cdot 2 - 2 \cdot 2 + 2 - 6 = -2.$$

Vidare, för $i = 1, 2, 3$ så är

$$V(X_i) = (D(X_i))^2 = 9$$

och

$$\begin{aligned} V(Y) &= V(3X_1 - 2X_2 + X_3 - 6) = V(3X_1 - 2X_2 + X_3) = \{\text{oberoende}\} \\ &= V(3X_1) + V(-2X_2) + V(X_3) = 3^2V(X_1) + (-2)^2V(X_2) + V(X_3) \\ &= 9 \cdot 9 + 4 \cdot 9 + 9 = 126 \end{aligned}$$

så

$$D(Y) = \sqrt{V(Y)} = \sqrt{126}.$$

5.23 Låt X beskriva längden av en typisk plank. Modell: X har väntevärde $E(X) = 2$ och $D(X) = \sigma = 5 \cdot 10^{-3}$. Då $n = 10$ plankor sågas till, låt Y beskriva plankornas sammanlagda längd om...

1.) ... alla plankorna sågas på samma gång. Då är $Y = 10X$ och

$$E(Y) = E(10X) = 10E(X) = 10 \cdot 2 = 20$$

och

$$V(Y) = V(10X) = 10^2V(X) = 100\sigma^2$$

och

$$D(Y) = \sqrt{V(Y)} = 10\sigma = 0.05.$$

2.) ... plankorna sågas till individuellt och får oberoende längder X_1, \dots, X_n . Då är $Y = \sum_{i=1}^{10} X_i$ och

$$E(Y) = E\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} E(X_i) = 10 \cdot 2 = 20$$

och

$$V(Y) = V\left(\sum_{i=1}^{10} X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{10} V(X_i) = 10\sigma^2$$

och

$$D(Y) = \sqrt{V(Y)} = \sqrt{10}\sigma \approx 0.0158.$$

b) Metoden i 2 ger mindre standardavvikelse för den totala längden Y .

5.24 Låt X , Y och Z vara oberoende stokastiska variabler sådana att

$$E(X + Y) = 1, \quad E(X - Z) = -4, \quad E(Y - Z) = -3.$$

Då är

$$\begin{aligned} 2E(X) &= E(2X) = E((X + Y) + (X - Z) - (Y - Z)) \\ &= E(X + Y) + E(X - Z) - E(Y - Z) = 1 - 4 + 3 = 0, \end{aligned}$$

det vill säga $E(X) = 0$. Detta medför att

$$E(Y) = E(Y) + E(X) = E(X + Y) = 1$$

och

$$E(Z) = E(Z) - E(X) = (-1)(E(X) - E(Z)) = (-1)E(X - Z) = (-1)(-4) = 4.$$

b) Då alla variabler är oberoende så är de automatisk okorrelerade varför

$$V(X + Y) = V(X) + V(Y) = 4 + 3 = 7 \neq 25,$$

$$V(X - Y) = V(X + (-1)Y) = V(X) + V((-1)Y) = V(X) + (-1)^2 V(Y) = 7.$$

Av samma anledning som $X - Y$ och $X + Y$ har samma varians måste $(X + Y) - Z$ och $(X + Y) + Z$ ha samma varians, varför

$$V(X \pm Y \pm Z) = V(X + Y + Z) = V(X) + V(Y) + V(Z) = 4 + 3 + 12 = 19.$$

Sålunda, det första påståendet b_1 är falskt, de två andra (b_2 och b_3) är sanna.

5.25 Låt X_1 och X_2 beskriva resultaten av två oberoende tärningskast. Då är $P(X_i = k) = 1/6$ för $k = 1, 2, \dots, 6$ och

$$E(X_i) = \sum_{k=1}^6 kP(X = k) = \frac{7}{2} = 3.5$$

och

$$V(X_i) = E(X^2) - (E(X))^2 = \sum_{k=1}^6 k^2 P(X = k) - (3.5)^2 = \frac{35}{12}$$

för $i = 1, 2$. Summan $X = X_1 + X_2$ har väntevärde

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 7$$

och varians

$$V(X) = V(X_1 + X_2) = \{\text{oberoende}\} = V(X_1) + V(X_2) = \frac{35}{6}.$$

För produkten $Y = X_1 X_2$ gäller

$$E(X_1 X_2) = \{\text{oberoende}\} = E(X_1) E(X_2) = \frac{49}{4}.$$

Variansen är lite krångligare, men upprepad användning av $V(Z) = E(Z^2) - (E(Z))^2$ ger

$$\begin{aligned} V(X_1 X_2) &= E((X_1 X_2)^2) - (E(X_1 X_2))^2 = E(X_1^2 X_2^2) - \left(\frac{49}{4}\right)^2 = \{\text{oberoende}\} \\ &= E(X_1^2) E(X_2^2) - \frac{2401}{16}. \end{aligned}$$

Men $E(X_1^2) = V(X_1) + (E(X_1))^2 = \frac{35}{12} + \left(\frac{7}{2}\right)^2 = \frac{91}{6}$ så

$$V(Y) = V(X_1 X_2) = \frac{91}{6} \cdot \frac{91}{6} - \frac{2401}{16} = \frac{11515}{144}.$$

Slutligen

$$C(X, Y) = E(XY) - E(X)E(Y) = E((X_1 + X_2)X_1 X_2) - E(X_1 + X_2)E(X_1 X_2)$$

där

$$\begin{aligned} E((X_1 + X_2)X_1 X_2) &= E(X_1^2 X_2 + X_2^2 X_1) = \{\text{oberoende}\} \\ &= E(X_1^2) E(X_2) + E(X_2^2) E(X_1) = \frac{91}{6} \cdot \frac{7}{2} + \frac{91}{6} \cdot \frac{7}{2} = \frac{637}{6}, \end{aligned}$$

dvs

$$C(X, Y) = E((X_1 + X_2)X_1 X_2) - E(X_1 + X_2)E(X_1 X_2) = \frac{637}{6} - 7 \cdot \frac{49}{4} = \frac{245}{12}.$$

Korrelationen fås till

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{245/12}{\sqrt{\frac{35}{6} \cdot \frac{11515}{144}}} = \sqrt{\frac{42}{47}}.$$

5.26 Vi låter X_A och X_B vara två stokastiska variabler sådana att

$$E(X_A) = E(X_B) = \mu \quad \text{och} \quad V(X_A) = \sigma_A^2, \quad V(X_B) = \sigma_B^2.$$

Med $Y = aX_A + bX_B$ så är

$$E(Y) = E(aX_A + bX_B) = aE(X_A) + bE(X_B) = a\mu + b\mu = (a + b)\mu$$

vilket ger $E(Y) = \mu$ om $a + b = 1$. Uttrycker vi a i b får vi att $a = 1 - b$ och $Y = (1 - b)X_A + bX_B$. Variansen för Y ges av

$$\begin{aligned} V(Y) &= V((1 - b)X_A + bX_B) = \{\text{oberoende}\} = (1 - b)^2 V(X_A) + b^2 V(X_B) \\ &= (1 - b)^2 \sigma_A^2 + b^2 \sigma_B^2. \end{aligned}$$

Denna minimeras med avseende på b genom lösning av

$$0 = \frac{d}{db} V(Y) = -2(1 - b)\sigma_A^2 + 2b\sigma_B^2 = 2[(\sigma_A^2 + \sigma_B^2)b - \sigma_A^2]$$

vilket ger

$$b = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2} \quad \text{och} \quad Y = (1 - b)X_A + bX_B = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2} X_A + \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2} X_B.$$

Kontroll av andraderivatans ger att detta verkligen är ett minimum.

5.27 Låt X_1, X_2, \dots, X_n beskriva personens vinst under $n = 12$ månader. Med modellen att

$$E(X_i) = -0.5, \quad D(X_i) = \sqrt{15}, \quad i = 1, 2, \dots, n$$

och att X_1, \dots, X_n är oberoende (okorrelerade räkter) är den totala vinsten under n månader $Y = X_1 + \dots + X_n$. Här är

$$E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = 12 \cdot (-0.5) = -6$$

och

$$V(Y) = V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(X_i) = 12 \cdot (\sqrt{15})^2 = 180.$$

5.28 Vi vet att korrelationen $\rho(X, Y) = C(X, Y) / \sqrt{V(X)V(Y)}$ uppfyller $-1 \leq \rho(X, Y) \leq 1$. Alltså är

$$-4 = -\sqrt{V(X)V(Y)} \leq C(X, Y) \leq \sqrt{V(X)V(Y)} = 4.$$

Detta medför att

$$V(X + Y) = \underbrace{V(X) + V(Y)}_{=17} + 2C(X, Y) = \begin{cases} \leq 17 + 2 \cdot 4 = 25 \\ \geq 17 - 2 \cdot 4 = 9. \end{cases}$$

5.29 För den stokastiska variabeln X är

$$E(X) = 0 \quad V(X) = 1 \quad E(X^3) = 2 \quad V(X^2) = 4$$

Alternativ 1: Då är

$$V(X + X^2) = V(X) + V(X^2) + 2C(X, X^2) = 1 + 4 + 2C(X, X^2) = 5 + 2C(X, X^2).$$

Nu är

$$E(X \cdot X^2) = E(X^3) = 2$$

och

$$E(X)E(X^2) = 0 \cdot E(X^2) = 0$$

varför $C(X, X^2) = E(X \cdot X^2) - E(X)E(X^2) = 2$ och således

$$V(X + X^2) = 5 + 2C(X, X^2) = 5 + 2 \cdot 2 = 9.$$

Alternativ 2: Vi bestämmer först $E(X^k)$, $k = 1, 2, 3, 4$ ur det givna.

$$\begin{aligned} E(X) &= 0 \\ E(X^2) &= V(X) + (E(X))^2 = 1 + 0 = 1 \\ E(X^3) &= 2 \\ E(X^4) &= E((X^2)^2) = V(X^2) + (E(X^2))^2 = 4 + 1 = 5 \end{aligned}$$

Med detta får vi att

$$\begin{aligned} V(X + X^2) &= E((X + X^2)^2) - (E(X + X^2))^2 = E(X^2 + 2X^3 + X^4) - (E(X) + E(X^2))^2 \\ &= E(X^2) + 2E(X^3) + E(X^4) - (0 + 1)^2 = 1 + 2 \cdot 2 + 5 - 1 = 9. \end{aligned}$$

5.30 Låt skolbarnens vikter beskrivas av de oberoende stokastiska variablerna X_1, \dots, X_n med väntevärde 36 och standardavvikelse 3. Det aritmetiska medelvärdet av n skolbarns vikter beskrivs av

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

där

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X) = 36 \text{ kg}$$

och

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{V(X)}{n}.$$

Alltså är

$$D(\bar{X}) = \sqrt{V(\bar{X})} = \frac{D(X)}{\sqrt{n}} = \frac{3}{\sqrt{3}} = \sqrt{3} \text{ kg}.$$

5.31 Låt X_1, \dots, X_n vara oberoende stokastiska variabler som beskriver en bestämning av smältpunkten för matfett där $D(X_i) = 2, i = 1, \dots, n$. Det aritmetiska medelvärdet av n smältpunktsbestämningar beskrivs av

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

där

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{V(X)}{n}$$

så

$$D(\bar{X}) = \sqrt{V(\bar{X})} = \frac{D(X)}{\sqrt{n}} = \frac{2}{\sqrt{n}}.$$

Kravet $D(\bar{X}) \leq 0.4$ ger $n \geq (2/0.4)^2 = 25$.

5.32 Beteckna $E(X) = \mu_x$ och $E(Y) = \mu_y$. Upprepad användning av $V(Z) = E(Z^2) - (E(Z))^2$ ger

$$\begin{aligned} V(XY) &= E((XY)^2) - (E(XY))^2 = E(X^2Y^2) - (E(XY))^2 = \{\text{oberoende}\} \\ &= E(X^2)E(Y^2) - (E(X)E(Y))^2 = E(X^2)E(Y^2) - \mu_x^2\mu_y^2 \\ &= (V(X) + \mu_x^2)(V(Y) + \mu_y^2) - \mu_x^2\mu_y^2 \\ &= V(X)V(Y) + V(X)\mu_y^2 + V(Y)\mu_x^2 + \mu_x^2\mu_y^2 - \mu_x^2\mu_y^2 \\ &= V(X)V(Y) + \mu_y^2V(X) + \mu_x^2V(Y). \end{aligned}$$

Generellt sett är $V(XY) \neq V(X)V(Y)$, men likhet gäller om $\mu_x = \mu_y = 0$.

5.33 Vi utnyttjar först att

$$\begin{aligned} C(X+Y, X-Y) &= C(X, X-Y) + C(Y, X-Y) \\ &= \underbrace{C(X, X)}_{=V(X)} - C(X, Y) + \underbrace{C(Y, X)}_{=C(X, Y)} - \underbrace{C(Y, Y)}_{V(Y)} = V(X) - V(Y). \end{aligned}$$

Nu är

$$\begin{aligned} V(X+Y)V(X-Y) &= (V(X) + V(Y) + 2C(X, Y))(V(X) + V(Y) - 2C(X, Y)) \\ &= (V(X))^2 + (V(Y))^2 + 2V(X)V(Y) = (V(X) + V(Y))^2 \end{aligned}$$

så

$$\rho(X+Y, X-Y) = \frac{C(X+Y, X-Y)}{\sqrt{V(X+Y)V(X-Y)}} = \frac{V(X) - V(Y)}{\sqrt{(V(X) + V(Y))^2}} = \frac{V(X) - V(Y)}{V(X) + V(Y)}.$$

5.34 Beteckna $V(X) = \sigma^2$. Först har vi att

$$C(\bar{X}_n, \bar{X}_n) = V(\bar{X}_n) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{okorrelerade}\} = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}.$$

Nu är

$$C(\bar{X}_n, \bar{X}_k) = C\left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{k} \sum_{j=1}^k X_j\right) = \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k C(X_i, X_j)$$

Om X_1, \dots, X_n är okorrelerade är $C(X_i, X_j) = 0$ för $j \neq i$ och

$$\frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k C(X_i, X_j) = \frac{1}{nk} \sum_{i=1}^k C(X_i, X_i) = \frac{1}{nk} \sum_{i=1}^k V(X_i) = \frac{1}{nk} \cdot k\sigma^2 = \frac{\sigma^2}{n}$$

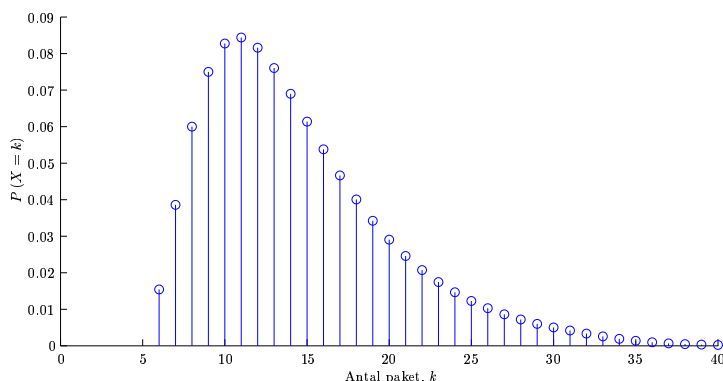
så

$$C(\bar{X}_n, \bar{X}_n - \bar{X}_k) = C(\bar{X}_n, \bar{X}_n) - C(\bar{X}_n, \bar{X}_k) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

vilket skulle visas.

5.35 För $i = 1, \dots, 6$, låt Y_i vara antalet paket man måste köpa för att öka samlingen från $i - 1$ till i djur. Möjliga värden Y_i är $\{1, 2, 3, \dots\}$. När man har $i - 1$ djur är sannolikheten $p_i = (n - (i - 1))/n$ att ett köpt paket innehåller ett nytt djur oberoende av tidigare paket. Alltså är Y_i för-första-gången-fördelad med parameter p_i vilket medför att $E(Y_i) = 1/p_i$. Nu är det totala antalet köpta paket $Y = Y_1 + Y_2 + \dots + Y_6$ där

$$\begin{aligned} E(Y) &= E(Y_1 + Y_2 + \dots + Y_6) = E(Y_1) + E(Y_2) + \dots + E(Y_6) = \frac{1}{p_1} + \frac{1}{p_2} + \dots + \frac{1}{p_6} \\ &= \frac{6}{6} + \frac{6}{6-1} + \dots + \frac{6}{1} = 6 \sum_{k=1}^6 \frac{1}{k} = 14.7. \end{aligned}$$



Fördelningen för antalet paket man är tvungen att köpa för att samla alla 6 plastdjur. Väntevärdet i fördelningen är 14.7.

5.36 Låt U_1, \dots, U_4 vara de positionerna för punkterna. Då är U_1, \dots, U_4 oberoende och likformigt fördelade på intervallet $[0, 1]$, det vill säga

$$f_U(x) = 1, \quad 0 \leq x \leq 1$$

vilket medför att

$$E(U) = \int_{-\infty}^{\infty} x f_U(x) dx = \int_0^1 x \cdot 1 dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

och

$$E(U^2) = \int_{-\infty}^{\infty} x^2 f_U(x) dx = \int_0^1 x^2 \cdot 1 dx = \left[\frac{x^3}{3} \right]_0^1 = \frac{1}{3}$$

så $V(U) = E(U^2) - (E(U))^2 = 1/3 - 1/4 = 1/12$. De fyra trianglarna har tillsammans area T där

$$T = \frac{(1 - U_4) \cdot U_1}{2} + \frac{(1 - U_1) \cdot U_2}{2} + \frac{(1 - U_2) \cdot U_3}{2} + \frac{(1 - U_3) \cdot U_4}{2}$$

Då $i \neq j$ är

$$E((1 - U_i)U_j) = E(U_j) - E(U_i U_j) = \{\text{oberoende}\} = E(U_j) - E(U_i)E(U_j) = \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

och

$$E((1 - U_i)U_i) = E(U_i) - E(U_i^2) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

så

$$E(T) = \frac{1}{2}E((1 - U_4)U_1) + \frac{1}{2}E((1 - U_1)U_2) + \frac{1}{2}E((1 - U_2)U_3) + \frac{1}{2}E((1 - U_3)U_4) = \frac{1}{2}.$$

Fyrhörningen har area $A = 1 - T$ så $E(A) = 1 - E(T) = 1/2$. För att bestämma $V(T)$ och $D(T)$ utnyttjas att $V(T) = E(T^2) - (E(T))^2$. Definiera $U_0 = U_4$ så kan man skriva

$$\begin{aligned} E((2T)^2) &= E\left(\left(\sum_{i=1}^4 (1 - U_{i-1})U_i\right)^2\right) = E\left(\sum_{i=1}^4 \sum_{j=1}^4 (1 - U_{i-1})U_i(1 - U_{j-1})U_j\right) \\ &= \sum_{i=1}^4 \sum_{j=1}^4 E((1 - U_{i-1})U_i(1 - U_{j-1})U_j). \end{aligned}$$

Vi skiljer på fallen $i = j$, $|i - j| = 1$ och $|i - j| > 1$.

- Då $i = j$ är

$$E((1 - U_{i-1})U_i(1 - U_{j-1})U_j) = E((1 - U_{i-1})^2 U_i^2) = E((1 - U_{i-1})^2) E(U_i^2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}.$$

- Då $i = j - 1$ är

$$\begin{aligned} E((1 - U_{i-1})U_i(1 - U_{j-1})U_j) &= E((1 - U_{i-1})U_i(1 - U_i)U_{i+1}) \\ &= E(1 - U_{i-1}) E(U_i(1 - U_i)) E(U_{i+1}) = \frac{1}{2} \cdot \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{24}. \end{aligned}$$

Samma väntevärde fås då $i - 1 = j$.

- Då $|i - j| > 1$ är

$$E((1 - U_{i-1})U_i(1 - U_{j-1})U_j) = E(1 - U_{i-1}) E(U_i) E(1 - U_{j-1}) E(U_j) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}.$$

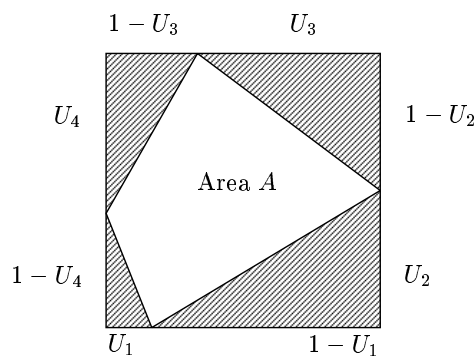
Sålunda fås

$$4E(T^2) = E((2T)^2) = \sum_{i=1}^4 \sum_{j=1}^4 E((1 - U_{i-1})U_i(1 - U_{j-1})U_j) = 4 \cdot \frac{1}{9} + 8 \cdot \frac{1}{24} + 4 \cdot \frac{1}{16} = \frac{37}{36}$$

och

$$V(T) = E(T^2) - (E(T))^2 = \frac{37}{4 \cdot 36} - \left(\frac{1}{2}\right)^2 = \frac{1}{4 \cdot 36}$$

så $D(T) = 1/12$.



Arean A bestäms som kvadratens area (1) minus triangelarnas totala area.

5.37 Låt Y_i , $i = 1, \dots, n$, vara stokastiska variabler sådana att

$$Y_i = \begin{cases} 1 & \text{om lapp med nummer } i \text{ dras i omgång } i \\ 0 & \text{annars.} \end{cases}$$

Då är $P(Y_i = 1) = 1/n$ och

$$E(Y_i) = \sum_k k \cdot P(Y_i = k) = 0 \cdot P(Y_i = 0) + 1 \cdot P(Y_i = 1) = \frac{1}{n}.$$

Med $Y = Y_1 + Y_2 + \dots + Y_n$ så är

$$E(Y) = E(Y_1 + Y_2 + \dots + Y_n) = E(Y_1) + E(Y_2) + \dots + E(Y_n) = n \cdot \frac{1}{n} = 1.$$

För att beräkna variansen måste beroendet mellan variablerna redas ut. Först beräknar vi

$$E(Y_i^2) = \sum_k k^2 \cdot P(Y_i = k) = 0^2 \cdot P(Y_i = 0) + 1^2 \cdot P(Y_i = 1) = \frac{1}{n}$$

och

$$C(Y_i, Y_i) = V(Y_i) = E(Y_i^2) - (E(Y_i))^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}.$$

Tag nu $i \neq j$. Då är

$$P(Y_i = 1, Y_j = 1) = P(Y_i = 1 | Y_j = 1) P(Y_j = 1) = \frac{1}{n-1} \cdot \frac{1}{n}.$$

Sålunda är

$$E(Y_i Y_j) = \sum_{(k,l)} kl \cdot P(Y_i = k, Y_j = l) = 1 \cdot 1 \cdot P(Y_i = 1, Y_j = 1) = \frac{1}{n(n-1)}.$$

Alltså är

$$C(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i) E(Y_j) = \frac{1}{n(n-1)} - \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2(n-1)}.$$

Nu är

$$\begin{aligned} V(Y) &= V\left(\sum_{i=1}^n Y_i\right) = C\left(\sum_{i=1}^n Y_i, \sum_{j=1}^n Y_j\right) = \sum_{i=1}^n \sum_{j=1}^n C(Y_i, Y_j) = \sum_{i=1}^n C(Y_i, Y_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n C(Y_i, Y_j) \\ &= n \cdot \frac{n-1}{n^2} + n(n-1) \cdot \frac{1}{n^2(n-1)} = 1 - \frac{1}{n} + \frac{1}{n} = 1 \end{aligned}$$

5.38 Låt $\mu_n = E(N)$ och $\mu_x = E(X)$. Genom att betinga på N får man att

$$E\left(\sum_{i=1}^N X_i \mid N = n\right) = E\left(\sum_{i=1}^n X_i \mid N = n\right) = \sum_{i=1}^n E(X_i | N = n) = \sum_{i=1}^n E(X_i) = n\mu_x$$

Så enligt lagen om total förväntan erhålles

$$\begin{aligned} E\left(\sum_{i=1}^N X_i\right) &= \sum_n E\left(\sum_{i=1}^N X_i \mid N = n\right) P(N = n) = \sum_n n\mu_x P(N = n) = \mu_x \underbrace{\sum_n n P(N = n)}_{E(N)} \\ &= \mu_x \mu_n. \end{aligned}$$

Låt $Y = \sum_{i=1}^N X_i$. Då är $V(Y) = E(Y^2) - (E(Y))^2$ där

$$E(Y^2) = E\left(\left(\sum_{i=1}^N X_i\right)^2\right) = E\left(\sum_{i=1}^N \sum_{j=1}^N X_i X_j\right)$$

Betingat $N = n$ är

$$E(Y^2|N = n) = \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j) = nE(X^2) + n(n-1)\mu_x^2 = nV(X) + n^2\mu_x^2$$

Alltså är enligt lagen om total förväntan

$$E(Y^2) = \sum_n E(Y^2|N = n)P(N = n) = \sum_n (nV(X) + n^2\mu_x^2)P(N = n) = V(X)\mu_n + \mu_x^2 E(N^2)$$

och således är

$$V(Y) = (V(X)\mu_n + \mu_x^2 E(N^2)) - (\mu_x \mu_n)^2 = V(X)\mu_n + \mu_x^2 V(N).$$

Med värden

$$\mu_x = 10, D(X) = 5 \quad \mu_n = 7, V(N) = 7$$

fås

$$E(Y) = \mu_n \mu_x = 70 \quad V(Y) = V(X)\mu_n + \mu_x^2 V(N) = 25 \cdot 7 + 100 \cdot 7 = 875,$$

det vill säga $D(Y) = \sqrt{875} \approx 29.58$.

6.1 Låt Z vara $N(0, 1)$ med fördelningsfunktion $\Phi(x) = P(Z \leq x)$. Då är

a) $P(Z \leq 1.82) = \Phi(1.82) = 0.9656$.

b) $P(Z \leq -0.35) = \Phi(-0.35) = 1 - \Phi(0.35) = 1 - 0.6368 = 0.3632$.

c) $P(-1.2 < Z < 0.5) = P(-1.2 < Z \leq 0.5) = \Phi(0.5) - \Phi(-1.2) = \Phi(0.5) - (1 - \Phi(1.2)) = 0.6915 - (1 - 0.8849) = 0.5764$.

d) a sådan att $0.05 = P(Z > a) = 1 - \Phi(a) = 1 - \Phi(\lambda_{0.05})$, vilket ger $a = \lambda_{0.05} = 1.6449$.

e) $a > 0$ sådan att $0.95 = P(|Z| < a) = 1 - (P(Z < -a) + P(Z > a)) = 1 - (\Phi(-a) + (1 - \Phi(a))) = 1 - (1 - \Phi(a) + 1 - \Phi(a)) = 2\Phi(a) - 1$. Vilket ger $\Phi(a) = 0.975$ eller $1 - \Phi(a) = 0.025 = 1 - \Phi(\lambda_{0.025})$ så $a = \lambda_{0.025} = 1.9600$.

6.2 Låt Z vara $N(0, 1)$ med fördelningsfunktion $\Phi(x) = P(Z \leq x)$. Då är

$$\begin{aligned} P(0.21 < X < 0.29) &= \Phi(0.29) - \Phi(0.21) = 0.61409 - 0.58317 = 0.030926 \\ P(-0.21 < X < 0.29) &= \Phi(0.29) - \Phi(-0.21) = \Phi(0.29) - [1 - \Phi(0.21)] = 0.61409 - 0.41683 \\ &= 0.19726 \\ P(-0.29 < X < -0.21) &= \Phi(-0.21) - \Phi(-0.29) = [1 - \Phi(0.21)] - [1 - \Phi(0.29)] \\ &= \Phi(0.29) - \Phi(0.21) = P(0.21 < X < 0.29) = 0.030926 \end{aligned}$$

6.3 Om X är $N(0, 1)$ så är $E(X) = 0$ och $D(X) = 1$, det vill säga $V(X) = 1$. Alltså är

$$E(Y) = E(3X + 2) = 3E(X) + 2 = 3 \cdot 0 + 2 = 2$$

och

$$V(Y) = V(3X + 2) = V(3X) = 3^2 V(X) = 9 \cdot 1 = 9$$

så $D(Y) = 3$. Notera att Y är $N(2, 3)$.

6.4 Låt X vara $N(\mu, \sigma)$ med $\mu = 5$ och $\sigma = 2$. Då är $Z = (X - \mu)/\sigma \sim N(0, 1)$ och

a)

$$P(X \leq 6) = P\left(\frac{X - \mu}{\sigma} \leq \frac{6 - \mu}{\sigma}\right) = P\left(Z \leq \frac{1}{2}\right) = \Phi(0.5) = 0.6915.$$

b)

$$\begin{aligned} P(1.8 < X < 7.2) &= P(1.8 < X \leq 7.2) = P\left(\frac{1.8 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{7.2 - \mu}{\sigma}\right) \\ &= P(-1.6 < Z \leq 1.1) = \Phi(1.1) - \Phi(-1.6) = \Phi(1.1) - (1 - \Phi(1.6)) \\ &= 0.80953. \end{aligned}$$

c) a sådan att $P(X \leq a) = 0.05$ bestäms ur

$$\begin{aligned} 1 - \Phi(\lambda_{0.05}) = 0.05 &= P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = P\left(Z \leq \frac{a - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{a - \mu}{\sigma}\right) = 1 - \Phi\left(-\frac{a - \mu}{\sigma}\right). \end{aligned}$$

Alltså är

$$-\frac{a - \mu}{\sigma} = \lambda_{0.05}$$

eller

$$a = \mu - \lambda_{0.05}\sigma = 5 - 1.6449 \cdot 2 = 1.7103.$$

6.5 Låt X vara $N(\mu, \sigma)$ med $\mu = -1$ och $\sigma = 0.01$. Då är $Z = (X - \mu)/\sigma \sim N(0, 1)$ och således

$$\begin{aligned} P(X < 0.99) &= P\left(\frac{X - \mu}{\sigma} < \frac{0.99 - \mu}{\sigma}\right) = P(Z \leq 199) = \Phi(199) \approx 1 \\ P(X < -0.99) &= P\left(\frac{X - \mu}{\sigma} < \frac{-0.99 - \mu}{\sigma}\right) = P(Z \leq 1) = \Phi(1) \approx 0.84134 \\ P(X > -0.99) &= 1 - P(X \leq -0.99) = 1 - P(X < -0.99) \approx 1 - 0.84134 = 0.15866 \\ P(-1.3 < X \leq -1.03) &= P\left(\frac{-1.3 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{-1.03 - \mu}{\sigma}\right) = P(-30 < Z \leq -3) \\ &= \Phi(-3) - \Phi(-30) = [1 - \Phi(3)] - [1 - \Phi(30)] = \Phi(30) - \Phi(3) \\ &\approx 1 - 0.99865 = 0.0013499. \end{aligned}$$

6.6 Låt X vara $N(\mu, \sigma)$ med $\mu = 20$ och $\sigma = 3$. Då är $(X - \mu)/\sigma \sim N(0, 1)$. Bestäm x så att

$$0.01 = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) = 1 - \Phi\left(-\frac{x - \mu}{\sigma}\right).$$

Kvantilen $\lambda_{0.01} = 2.3263$ är sådan att $1 - \Phi(\lambda_{0.01}) = 0.01$ det vill säga

$$\lambda_{0.01} = -\frac{x - \mu}{\sigma}$$

eller

$$x = \mu - \lambda_{0.01}\sigma = 20 - 2.3263 \cdot 3 = 13.021.$$

6.7 Låt X vara $N(\mu, \sigma)$ med $\mu = 180$ och $\sigma = 5$. Då är $Z = (X - \mu)/\sigma \sim N(0, 1)$. Då är

$$\begin{aligned} P(X \geq 170) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{170 - \mu}{\sigma}\right) = P(Z \geq -2) = 1 - P(Z \leq -2) = 1 - \Phi(-2) \\ &= 1 - [1 - \Phi(2)] = \Phi(2) = 0.97725. \end{aligned}$$

Vidare så är

$$\begin{aligned} P(170 \leq X \leq 200) &= P\left(\frac{170 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{200 - \mu}{\sigma}\right) = P(-2 \leq Z \leq 4) = \Phi(4) - \Phi(-2) \\ &= \Phi(4) - [1 - \Phi(2)] \approx 0.99997 - 1 + 0.97725 = 0.97722. \end{aligned}$$

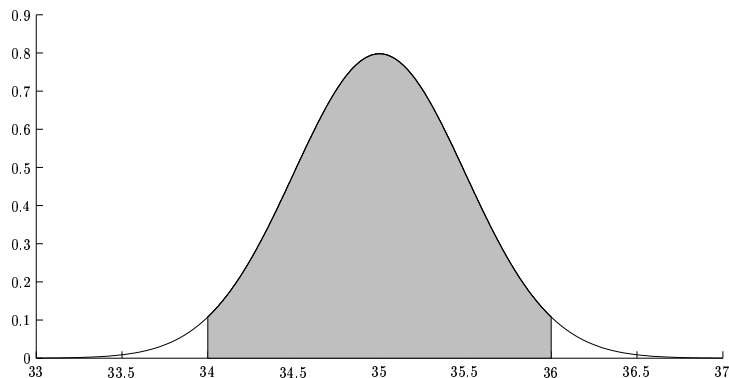
6.8 Låt X beskriva mängden kaffe i säcken, X är $N(\mu, \sigma)$ där $\mu = 35$ och $\sigma = 0.5$.

a) Då är sannolikheten att säcken räcker till minst 36 burkar

$$\begin{aligned} P(X \geq 36) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{36 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{36 - \mu}{\sigma}\right) \\ &= 1 - \Phi(2) = 1 - 0.97725 = 0.02275. \end{aligned}$$

b) Då är sannolikheten att säcken räcker till 34 men inte till 36 burkar

$$\begin{aligned} P(34 \leq X < 36) &= P\left(\frac{34 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} < \frac{36 - \mu}{\sigma}\right) = \Phi(2) - \Phi(-2) \\ &= \Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1 = 2 \cdot 0.97725 - 1 = 0.9545. \end{aligned}$$



Tätheten $f_X(x)$ för $N(35, 0.5)$.

6.9 Låt X vara $N(\mu, \sigma)$ med $\mu = 0.5\text{kg}$ och $\sigma = 0.003\text{kg}$. Då är $Z = (X - \mu)/\sigma \sim N(0, 1)$ och

$$\begin{aligned} P(X \geq 0.495) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{0.495 - \mu}{\sigma}\right) = P\left(Z \geq -\frac{5}{3}\right) = 1 - \Phi\left(-\frac{5}{3}\right) = 1 - \left[1 - \Phi\left(\frac{5}{3}\right)\right] \\ &= \Phi(5/3) = 0.95221. \end{aligned}$$

Vi söker nu d så att

$$P(0.5 - d \leq X \leq 0.5 + d) = 1 - \alpha.$$

Då är

$$\begin{aligned} 1 - \alpha &= P(0.5 - d \leq X \leq 0.5 + d) = P\left(\frac{0.5 - d - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{0.5 + d - \mu}{\sigma}\right) \\ &= P\left(-\frac{d}{\sigma} \leq Z \leq \frac{d}{\sigma}\right) = \Phi\left(\frac{d}{\sigma}\right) - \Phi\left(-\frac{d}{\sigma}\right) = 2\Phi\left(\frac{d}{\sigma}\right) - 1 \end{aligned}$$

eller $1 - \Phi(d/\sigma) = \alpha/2$. Eftersom $1 - \Phi(\lambda_{\alpha/2}) = \alpha/2$ är $d/\sigma = \lambda_{\alpha/2}$ eller $d = \lambda_{\alpha/2}\sigma$. Sålunda erhålles

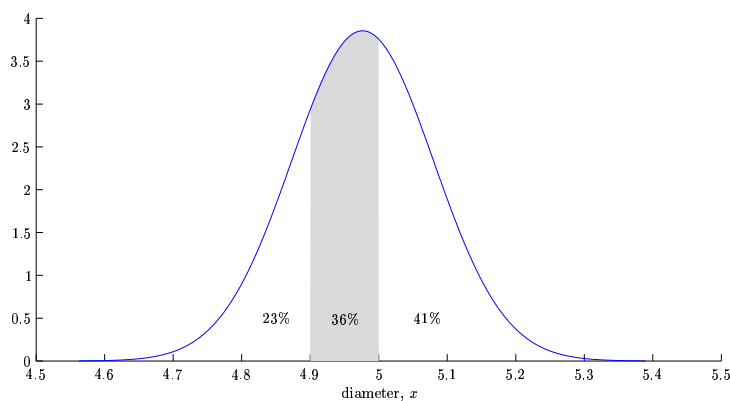
$1 - \alpha$	$\lambda_{\alpha/2}$	d [kg]
0.50	0.67449	$2.0235 \cdot 10^{-3}$
0.95	1.9600	$5.8799 \cdot 10^{-3}$
0.99	2.5758	$7.7275 \cdot 10^{-3}$

6.10 Låt X beskriva en kulas diameter. Med modellen att X är $N(\mu, \sigma)$ så kan de givna antagandena formuleras

$$\begin{aligned} 0.23 &= P(X \leq 4.9) = P\left(\frac{X - \mu}{\sigma} \leq \frac{4.9 - \mu}{\sigma}\right) = \Phi\left(\frac{4.9 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu - 4.9}{\sigma}\right) = 1 - \Phi(\lambda_{0.23}) \\ 0.59 &= P(X \leq 5.0) = P\left(\frac{X - \mu}{\sigma} \leq \frac{5.0 - \mu}{\sigma}\right) = \Phi\left(\frac{5.0 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu - 5.0}{\sigma}\right) = 1 - \Phi(\lambda_{0.59}) \end{aligned}$$

Ur tabell får man att $\lambda_{0.23} = 0.7388$ och $\lambda_{0.59} = -\lambda_{1-0.59} = -\lambda_{0.41} = -0.2275$. Alltså är

$$\begin{cases} \frac{\mu - 4.9}{\sigma} = \lambda_{0.23} \\ \frac{\mu - 5.0}{\sigma} = \lambda_{0.59} \end{cases} \quad \text{vilket ger} \quad \begin{cases} \mu = \frac{9.9}{2} + \frac{0.1}{2} \frac{\lambda_{0.23} + \lambda_{0.59}}{\lambda_{0.23} - \lambda_{0.59}} = 4.9765 \\ \sigma = \frac{0.1}{\lambda_{0.23} - \lambda_{0.59}} = 0.1035 \end{cases}$$



Täthetsfunktionen för kullagrenas diametrar.

6.11 Låt X beskriva diametern för en svarvad cylinder och ansätt modellen att X är $N(\mu, \sigma) = N(12.4, \sigma)$. Vi vill bestämma σ så att

$$\begin{aligned} 0.95 &= P(12.0 \leq X \leq 12.8) = P\left(\frac{12.0 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{12.8 - \mu}{\sigma}\right) = \Phi\left(\frac{0.4}{\sigma}\right) - \Phi\left(-\frac{0.4}{\sigma}\right) \\ &= 2\Phi\left(\frac{0.4}{\sigma}\right) - 1, \end{aligned}$$

med andra ord, så att $1 - \Phi(0.4/\sigma) = 0.025 = 1 - \Phi(\lambda_{0.025})$. Alltså är $\sigma = 0.4/\lambda_{0.025} = 0.4/1.96 = 0.2041$.

6.12 Låt X beskriva vikten av en typisk person. Modell: X är $N(70, 10)$. Låt Y beskriva vikten för $n = 10$ personer,

$$Y = X_1 + X_2 + \cdots + X_n,$$

där X_1, \dots, X_n är oberoende och fördelade som X . Då är Y normalfördelad med väntevärde

$$\mu = E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X) = 10 \cdot 70 = 700$$

och varians

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(X_i) = nV(X) = 10 \cdot 10^2 = 1000$$

dvs Y är $N(\mu, \sigma) = N(700, \sqrt{1000})$. Sökt är

$$P(Y > 800) = P\left(\frac{Y - \mu}{\sigma} > \frac{800 - \mu}{\sigma}\right) = 1 - \Phi(\sqrt{10}) = 0.0007827.$$

6.13 Låt X och Y vara oberoende där X är $N(1, 1)$ och Y är $N(-1, 2)$.

a) Då är...

- ... $X + Y$ normalfördelad med väntevärde $E(X + Y) = E(X) + E(Y) = 1 + (-1) = 0$ och varians $V(X + Y) = \{\text{oberoende}\} = V(X) + V(Y) = 1^2 + 2^2 = 5$, det vill säga $D(X + Y) = \sqrt{5}$. Alltså $X + Y \sim N(0, \sqrt{5})$.
- ... $X - Y$ normalfördelad med väntevärde $E(X - Y) = E(X) - E(Y) = 1 - (-1) = 2$ och varians $V(X - Y) = \{\text{oberoende}\} = V(X) + (-1)^2 V(Y) = 1^2 + 2^2 = 5$, det vill säga $D(X - Y) = \sqrt{5}$. Alltså $X - Y \sim N(2, \sqrt{5})$.

Notera att när X och Y är oberoende (okorrelerade) så har $X + Y$ och $X - Y$ samma varians.

6.14 Låt X_1 och X_2 beskriva tiderna i byggprojektets två faser. Modell: X_1 och X_2 är oberoende och X_1 är $N(200, 20)$ och X_2 är $N(100, 15)$. Den totala byggtiden $Y = X_1 + X_2$ är då normalfördelad med väntevärde

$$\mu = E(Y) = E(X_1 + X_2) = E(X_1) + E(X_2) = 200 + 100 = 300$$

och varians

$$\sigma^2 = V(Y) = V(X_1 + X_2) = \{\text{oberoende}\} = V(X_1) + V(X_2) = 20^2 + 15^2 = 625$$

dvs Y är $N(\mu, \sigma) = N(300, 25)$. Sökt är

$$P(Y > 310) = P\left(\frac{Y - \mu}{\sigma} > \frac{310 - \mu}{\sigma}\right) = 1 - \Phi(0.4) = 0.34458.$$

6.15 Låt X beskriva tiden till relä 1 löser ut och Y den för relä 2. Modell: X är $N(1, 0.1)$ och Y är $N(1.5, 0.2)$ där X och Y är oberoende stokastiska variabler. Relä 2 löser ut före relä 1 om $Y < X$ eller, omformulerat, då $Y - X < 0$. Med $W = Y - X$ så är W normalfördelad med väntevärde

$$\mu = E(W) = E(Y - X) = E(Y) - E(X) = 1.5 - 1 = 0.5$$

och varians

$$\sigma^2 = V(W) = V(Y - X) = \{\text{oberoende}\} = V(Y) + (-1)^2 V(X) = 0.2^2 + 0.1^2 = 0.05$$

dvs W är $N(0.5, \sqrt{0.05})$. Alltså,

$$\begin{aligned} P(Y - X < 0) &= P(W < 0) = P\left(\frac{W - \mu}{\sigma} < \frac{0 - \mu}{\sigma}\right) = \Phi(-\mu/\sigma) = \Phi(-\sqrt{5}) \\ &= 1 - \Phi(\sqrt{5}) = 1 - 0.9873 = 0.0127. \end{aligned}$$

6.16 Låt X och Y vara oberoende där X är $N(150, 3)$ och Y är $N(100, 4)$.

a) Då är...

- ... $X + Y$ normalfördelad med väntevärde $E(X + Y) = E(X) + E(Y) = 150 + 100 = 250$ och varians $V(X + Y) = \{\text{oberoende}\} = V(X) + V(Y) = 3^2 + 4^2 = 25$, det vill säga $D(X + Y) = 5$. Alltså $X + Y \sim N(250, 5)$.
- ... $X - Y$ normalfördelad med väntevärde $E(X - Y) = E(X) - E(Y) = 150 - 100 = 50$ och varians $V(X - Y) = \{\text{oberoende}\} = V(X) + (-1)^2 V(Y) = 3^2 + 4^2 = 25$, det vill säga $D(X - Y) = 5$. Alltså $X - Y \sim N(50, 5)$.
- ... $(X + Y)/2$ normalfördelad med väntevärde $E((X + Y)/2) = \frac{1}{2}E(X + Y) = 125$ och varians $V(\frac{1}{2}(X + Y)) = (\frac{1}{2})^2 V(X + Y) = \frac{25}{4}$, det vill säga $D((X + Y)/2) = 2.5$. Alltså $(X + Y)/2 \in N(125, 2.5)$.

b) Låt Z vara $N(0, 1)$. Nu är

$$\begin{aligned} P(X + Y < 242.6) &= P\left(\frac{(X + Y) - 250}{5} < \frac{242.6 - 250}{5}\right) = P(Z < -1.48) = \Phi(-1.48) \\ &= 1 - \Phi(1.48) \approx 1 - 0.9306 = 0.0694, \end{aligned}$$

$$\begin{aligned} P(|X - Y| < 40) &= P(-40 < X - Y < 40) = P\left(\frac{-40 - 50}{5} < \frac{(X - Y) - 50}{5} < \frac{40 - 50}{5}\right) \\ &= P(-18 < Z < -2) = \Phi(-2) - \Phi(-18) \\ &= [1 - \Phi(2)] - [1 - \Phi(18)] \approx 0.02275 - 0 = 0.02275 \end{aligned}$$

$$\begin{aligned} P\left(\left|\frac{X + Y}{2} - 125\right| > 5\right) &= 1 - P\left(-5 \leq \frac{X + Y}{2} - 125 \leq 5\right) = 1 - P\left(\frac{-5}{2.5} \leq \frac{\frac{X + Y}{2} - 125}{2.5} \leq \frac{5}{2.5}\right) \\ &= 1 - P(-2 \leq Z \leq 2) = 1 - (\Phi(2) - \Phi(-2)) = 1 - (\Phi(2) - [1 - \Phi(2)]) \\ &= 2(1 - \Phi(2)) \approx 2(1 - 0.9772) = 0.0455. \end{aligned}$$

6.17 Låt X_1 och X_2 vara oberoende och $N(\mu, \sigma) = N(1, 2)$. Då är $\bar{X} = (X_1 + X_2)/2$ normalfördelad med väntevärde

$$E(\bar{X}) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}\underbrace{E(X_1)}_{=\mu} + \frac{1}{2}\underbrace{E(X_2)}_{=\mu} = \mu = 1$$

och varians

$$V(\bar{X}) = V\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \{\text{oberoende}\} = \left(\frac{1}{2}\right)^2 \underbrace{V(X_1)}_{=\sigma^2} + \left(\frac{1}{2}\right)^2 \underbrace{V(X_2)}_{=\sigma^2} = \frac{\sigma^2}{2} = 2$$

det vill säga $D(\bar{X}) = \sqrt{2}$. Alltså, \bar{X} är $N(1, \sqrt{2})$.

6.18 Låt X_1, \dots, X_n vara oberoende och $N(\mu, \sigma) = N(\mu, 0.2)$. Med $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ är $\bar{X} - \mu = \frac{1}{n} \sum_{i=1}^n X_i - \mu$ normalfördelad med väntevärde

$$E(\bar{X} - \mu) = E\left(\frac{1}{n} \sum_{i=1}^n X_i - \mu\right) = \frac{1}{n} \sum_{i=1}^n \underbrace{E(X_i)}_{=\mu} - \mu = \frac{1}{n} \cdot n \cdot \mu - \mu = 0$$

och varians

$$V(\bar{X} - \mu) = V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \underbrace{V(X_i)}_{=\sigma^2} = \frac{\sigma^2}{n}$$

det vill säga $D(\bar{X} - \mu) = \sigma/\sqrt{n} = 0.2/\sqrt{n}$.

a) Alltså, $\bar{X} - \mu$ är $N(0, 0.2/\sqrt{n})$.

b) Med normalfördelningen för $\bar{X} - \mu$ så är $Z = (\bar{X} - \mu)/(0.2/\sqrt{n})$ $N(0, 1)$ -fördelad och

$$\begin{aligned} P(|\bar{X} - \mu| > 0.2/\sqrt{n}) &= P\left(\left|\frac{\bar{X} - \mu}{0.2/\sqrt{n}}\right| > 1\right) = P(|Z| > 1) = P(Z > 1) + P(Z < -1) \\ &= 1 - \Phi(1) + \Phi(-1) = 2(1 - \Phi(1)) = 2(1 - 0.8413) = 0.3173. \end{aligned}$$

c) Med $n = 16$ så är

$$\begin{aligned} P(|\bar{X} - \mu| > 0.1) &= P\left(\left|\frac{\bar{X} - \mu}{0.2/\sqrt{n}}\right| > \frac{0.1}{0.2/\sqrt{n}}\right) = P(|Z| > 2) = P(Z > 2) + P(Z < -2) \\ &= 1 - \Phi(2) + \Phi(-2) = 2(1 - \Phi(2)) = 2(1 - 0.9773) = 0.0455. \end{aligned}$$

d) Nu är

$$P(|\bar{X} - \mu| > 0.01) = P\left(\left|\frac{\bar{X} - \mu}{0.2/\sqrt{n}}\right| > \frac{0.01}{0.2/\sqrt{n}}\right) = 2\left(1 - \Phi\left(\frac{0.01}{0.2/\sqrt{n}}\right)\right) \leq 0.001$$

så

$$1 - \Phi\left(\frac{0.01}{0.2/\sqrt{n}}\right) \leq 0.0005 = 1 - \Phi(\lambda_{0.0005})$$

vilket ger

$$\frac{0.01}{0.2/\sqrt{n}} \geq \lambda_{0.0005}$$

eller

$$n \geq \left(\frac{0.2\lambda_{0.0005}}{0.01}\right)^2 = \left(\frac{0.2 \cdot 3.2905}{0.01}\right)^2 = 4331.$$

6.19 Låt X beskriva vikten av en tablett. Ansätt modellen att X är $N(\mu, \sigma)$ = $N(0.65, 0.02)$.

a) Då är

$$\begin{aligned} P(X \notin (0.60, 0.70)) &= 1 - P(0.60 \leq X \leq 0.70) = 1 - P\left(\frac{0.60 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{0.70 - \mu}{\sigma}\right) \\ &= 1 - (\Phi(2.5) - \Phi(-2.5)) = 2(1 - \Phi(2.5)) = 2(1 - 0.9938) = 0.0124. \end{aligned}$$

b) Baserat på $n = 30$ oberoende observationer X_1, \dots, X_n så är $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ normalfördelad, \bar{X} är $N(\mu, \sigma/\sqrt{n})$. Alltså är

$$\begin{aligned} P(\bar{X} \notin (0.64, 0.66)) &= 1 - P(0.64 \leq \bar{X} \leq 0.66) = 1 - P\left(\frac{0.64 - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.66 - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - (\Phi(2.7386) - \Phi(-2.7386)) = 2(1 - \Phi(2.7386)) = 2(1 - 0.9969) \\ &= 0.0062. \end{aligned}$$

c) Nu är

$$P(|\bar{X} - \mu| > 0.01) = P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| > \frac{0.01}{\sigma/\sqrt{n}}\right) = 2\left(1 - \Phi\left(\frac{0.01}{\sigma/\sqrt{n}}\right)\right) \leq 0.01$$

så

$$1 - \Phi\left(\frac{0.01}{\sigma/\sqrt{n}}\right) \leq 0.005 = 1 - \Phi(\lambda_{0.005})$$

vilket ger

$$\frac{0.01}{\sigma/\sqrt{n}} \geq \lambda_{0.005}$$

eller

$$n \geq \left(\frac{\sigma \lambda_{0.005}}{0.01} \right)^2 = \left(\frac{0.02 \cdot 2.5758}{0.01} \right)^2 = 26.54,$$

det vill säga 27 eller fler tabletter.

6.20 Låt X beskriva vikten i ton av en typisk järnvägsvagn. Modell: X har $E(X) = 10$ och $D(X) = 0.5$. Låt Y beskriva vikten för $n = 25$ vagnar,

$$Y = X_1 + X_2 + \dots + X_n,$$

där X_1, \dots, X_n är oberoende och fördelade som X . Då har Y väntevärde

$$\mu = E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X) = 25 \cdot 10 = 250$$

och varians

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(X_i) = nV(X) = 25 \cdot (0.5)^2 = 25/4$$

dvs $D(Y) = 5/2$ ton. Enligt Centrala gränsvärdessatsen är Y approximativt $N(\mu, \sigma) = N(250, 2.5)$. Sökt är

$$P(Y > 255) = P\left(\frac{Y - \mu}{\sigma} > \frac{255 - \mu}{\sigma}\right) \approx 1 - \Phi(2) = 0.02275.$$

6.21 Låt X beskriva antalet barn i förskoleålder i ett hushåll. Modell: X har sannolikhetsfunktion

$$p_X(0) = 0.40 \quad p_X(1) = 0.20 \quad p_X(2) = 0.30 \quad p_X(3) = 0.10.$$

Då är

$$E(X) = \sum_k kP(X = k) = 0 \cdot 0.40 + \dots + 3 \cdot 0.10 = 1.1$$

och

$$V(X) = E((X - E(X))^2) = \sum_k (k - 1.1)^2 P(X = k) = 1.09.$$

Låt Y beskriva antalet barn i förskoleåldern i ett område med $n = 1000$ hushåll,

$$Y = X_1 + X_2 + \dots + X_n,$$

där X_1, \dots, X_n är oberoende och fördelade som X . Då har Y väntevärde

$$\mu = E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X) = 1000 \cdot 1.1 = 1100$$

och varians

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(X_i) = nV(X) = 1000 \cdot 1.09 = 1090.$$

dvs $D(Y) = \sqrt{1090}$ barn. Enligt CGS är Y approximativt $N(\mu, \sigma) = N(1100, \sqrt{1090})$. Vi söker y så att $P(Y < y) = 0.90$.

$$1 - \Phi(\lambda_{0.10}) = 0.10 = P(Y > y) = P\left(\frac{Y - \mu}{\sigma} > \frac{y - \mu}{\sigma}\right) \approx 1 - \Phi\left(\frac{y - \mu}{\sigma}\right).$$

Alltså är approximativt

$$\frac{y - \mu}{\sigma} = \lambda_{0.10}$$

eller

$$y = \mu + \lambda_{0.10}\sigma = 1100 + 1.2816 \cdot \sqrt{1090} = 1142.3.$$

Alltså, bygg 1143 platser.

6.22 Låt X_1, \dots, X_n vara oberoende och likformigt fördelade på $[-0.5, 0.5]$. Då är

$$E(X_i) = 0 \quad \text{och} \quad V(X_i) = \frac{1}{12}.$$

Med $Y = \sum_{i=1}^n X_i$ så har Y väntevärde

$$\mu = E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X) = 1000 \cdot 0 = 0$$

och varians

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(X_i) = \frac{n}{12}.$$

Enligt Centrala gränsvärdesatsen är Y för stora n approximativt $N(\mu, \sigma) = N(0, \sqrt{n/12})$. Alltså är

$$\begin{aligned} P(|Y| < K\sqrt{n}) &= P(-K\sqrt{n} < Y < K\sqrt{n}) \\ &= P\left(\frac{-K\sqrt{n}}{\sqrt{n/12}} < \frac{Y-0}{\sqrt{n/12}} < \frac{K\sqrt{n}}{\sqrt{n/12}}\right) \approx \Phi(K\sqrt{12}) - \Phi(-K\sqrt{12}) = 2\Phi(K\sqrt{12}) - 1. \end{aligned}$$

Det sista ledet är oberoende av n så felet $|Y - 0|$ växer i storleksordningen \sqrt{n} . Notera att n påverkar dock precisionen hos normalapproximationen.

För olika värden på K fås sannolikheterna

$$2\Phi(K\sqrt{12}) - 1 = \begin{cases} 0.27097 & \text{om } K = 0.1 \\ 0.91674 & \text{om } K = 0.5 \\ 0.99947 & \text{om } K = 1. \end{cases}$$

6.23 Låt X_1, X_2, \dots, X_n , $n = 50$, beskriva livslängderna för elektronrören. Modell: X_i är oberoende och exponentialfördelade med väntevärde 200 (intensitet $1/200$). Den totala tiden som lagret räcker beskrivs av $T = \sum_{i=1}^n X_i$ som har väntevärde

$$\mu = E(T) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = 50 \cdot 200 = 10000$$

och varians

$$\sigma^2 = V(T) = V\left(\sum_{i=1}^n X_i\right) = \{\text{ober.}\} = \sum_{i=1}^n \underbrace{V(X_i)}_{=200^2} = 50 \cdot 200^2 = 2 \cdot 10^6.$$

Vi söker tiden t sådan att $P(T > t) = 0.90$. Enligt Centrala gränsvärdesatsen är T approximativt normalfördelad och

$$0.90 = P(T > t) = P\left(\frac{T - \mu}{\sigma} > \frac{t - \mu}{\sigma}\right) \approx 1 - \Phi\left(\frac{t - \mu}{\sigma}\right),$$

dvs $(t - \mu)/\sigma \approx \lambda_{0.90} = -\lambda_{0.10} = -1.2816$, eller

$$t \approx \mu - \lambda_{0.10}\sigma = 10000 - 1.2816 \cdot 1414.2 = 8187.6 \text{ timmar.}$$

6.24 Låt X_1, \dots, X_n vara oberoende och likafördelade med $E(X_i) = 0.1$ och $D(X_i) = 8$. Då är $\sum_{i=1}^n X_i$ approximativt normalfördelad för stora n och således även $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Alltså, \bar{X} är approximativt $N(\mu, \sigma)$ där

$$\mu = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = 0.1$$

och

$$\sigma^2 = V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{8^2}{n}$$

det vill säga $\sigma = 8/\sqrt{n}$.

Nu är

$$\begin{aligned} P(|\bar{X}| > 0.08) &= 1 - P(-0.08 \leq \bar{X} \leq 0.08) = 1 - P\left(\frac{-0.08 - 0.1}{8/\sqrt{n}} \leq \frac{\bar{X} - 0.1}{8/\sqrt{n}} \leq \frac{0.08 - 0.1}{8/\sqrt{n}}\right) \\ &= 1 - \left(\Phi\left(-\frac{0.02}{8/\sqrt{n}}\right) - \Phi\left(-\frac{0.18}{8/\sqrt{n}}\right)\right) = 1 - \Phi\left(\frac{0.18}{8/\sqrt{n}}\right) + \Phi\left(\frac{0.02}{8/\sqrt{n}}\right) \end{aligned}$$

så

a) med $n = 16$ erhålls $1 - \Phi(0.09) + \Phi(0.01) = 1 - 0.53586 + 0.50399 = 0.96814$.

b) med $n = 1600$ erhålls $1 - \Phi(0.9) + \Phi(0.1) = 1 - 0.81594 + 0.53983 = 0.72389$.

c) med $n = 160\,000$ erhålls $1 - \Phi(9) + \Phi(1) = 1 - 1 + 0.84134 = 0.84134$.

6.25 Låt X beskriva vikten för en magnecyltablett. Modell: X är en stokastisk variabel med $E(X) = 0.65$ och $D(X) = 0.02$ [gram].

a) Låt Y beskriva den sammanlagda vikten av $n = 100$ tabletter. Då är $Y = X_1 + \dots + X_n$ där X_1, \dots, X_n är oberoende och fördelade som X . Vidare så är

$$E(Y) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = 0.65 + \dots + 0.65 = 65 \text{ gram}$$

och

$$V(Y) = V(X_1 + \dots + X_n) = \{\text{oberoende}\} = V(X_1) + \dots + V(X_n) = 0.02^2 + \dots + 0.02^2 = 0.04 \text{ gram}^2,$$

det vill säga $D(Y) = 0.2$ gram.

b) Enligt Centrala gränsvärdessatsen är Y approximativt normalfördelad med väntevärde $\mu = 65$ och standardavvikelse $\sigma = 0.2$. Alltså är

$$P(Y \leq 65.3) = P\left(\frac{Y - \mu}{\sigma} \leq \frac{65.3 - \mu}{\sigma}\right) \approx \Phi\left(\frac{65.3 - \mu}{\sigma}\right) = \Phi(1.50) = 0.9332.$$

6.26 Låt X beskriva vikten för en magnecyltablett. Modell: X är en stokastisk variabel med $E(X) = 0.65$ och $D(X) = 0.02$ [gram].

Låt Y beskriva den sammanlagda vikten av $n = 99$ tabletter. Då är $Y = X_1 + \dots + X_n$ där X_1, \dots, X_n är oberoende och fördelade som X . Vidare så är

$$\mu = E(Y) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = 0.65 + \dots + 0.65 = 64.35 \text{ gram}$$

och

$$V(Y) = V(X_1 + \dots + X_n) = \{\text{oberoende}\} = V(X_1) + \dots + V(X_n) = 0.02^2 + \dots + 0.02^2 = 0.0396 \text{ gram}^2,$$

det vill säga $\sigma = D(Y) = 0.1990$. Enligt Centrala gränsvärdesatsen är Y approximativt normalfördelad med väntevärde μ och standardavvikelse σ .

Alltså asken innehåller minst 100 tabletter om

$$P(Y < 65) = P\left(\frac{Y - \mu}{\sigma} < \frac{65 - \mu}{\sigma}\right) \approx \Phi\left(\frac{65 - \mu}{\sigma}\right) = \Phi(3.266) \approx 0.9995.$$

6.27 Låt Z vara en $N(0, 1)$ -fördelad stokastisk variabel. Då är $X = \mu + \sigma Z$ normalfördelad med väntevärde

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma \underbrace{E(Z)}_{=0} = \mu$$

och varians

$$V(X) = V(\mu + \sigma Z) = \sigma^2 \underbrace{V(Z)}_{=1} = \sigma^2$$

det vill säga standardavvikelse σ om $\sigma > 0$.

Låt Z_1 och Z_2 vara oberoende $N(0, 1)$ -fördelade stokastiska variabler.

Då är $Y = \sigma_Y(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) + \mu_Y$ normalfördelad med väntevärde

$$E(Y) = E(\sigma_Y(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) + \mu_Y) = \sigma_Y(\rho E(Z_1) + \sqrt{1 - \rho^2} E(Z_2)) + \mu_Y = \mu_Y$$

och varians

$$\begin{aligned} V(Y) &= V(\sigma_Y(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) + \mu_Y) = \{\text{oberoende}\} = \sigma_Y^2 (\rho^2 V(Z_1) + (1 - \rho^2)V(Z_2)) \\ &= \sigma_Y^2 (\rho^2 + 1 - \rho^2) = \sigma_Y^2. \end{aligned}$$

Slutligen, kovariansen mellan X och Y

$$\begin{aligned} C(X, Y) &= C(\mu_X + \sigma_X Z_1, \sigma_Y(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) + \mu_Y) = C(\sigma_X Z_1, \sigma_Y(\rho Z_1 + \sqrt{1 - \rho^2} Z_2)) \\ &= \sigma_X \sigma_Y \rho \underbrace{C(Z_1, Z_1)}_{=V(Z_1)=1} + \sigma_X \sigma_Y \sqrt{1 - \rho^2} \underbrace{C(Z_1, Z_2)}_{=0} = \sigma_X \sigma_Y \rho \end{aligned}$$

så korrelationen mellan X och Y är

$$\frac{C(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_X \sigma_Y \rho}{\sigma_X \sigma_Y} = \rho.$$

6.28 Låt (X, Y) beskriva skillnaden mellan punktens uttryckade och faktiska position där X och Y är oberoende $N(0, 1)$ -fördelade stokastiska variabler. Avståndet mellan uttryckt och faktisk position ges av $R = \sqrt{X^2 + Y^2}$ och vi söker $P(R \leq t)$ då $t = 2$. För $t > 0$ är

$$P(R \leq t) = P(\sqrt{X^2 + Y^2} \leq t) = P(X^2 + Y^2 \leq t^2) = \int \int_{(x,y): x^2+y^2 \leq t^2} f_{X,Y}(x, y) dx dy$$

utnyttja oberoendet

$$\begin{aligned} &= \int \int_{(x,y): x^2+y^2 \leq t^2} f_X(x) f_Y(y) dx dy = \int \int_{(x,y): x^2+y^2 \leq t^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dx dy \\ &= \int \int_{(x,y): x^2+y^2 \leq t^2} \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy \end{aligned}$$

inför polära koordinater r och θ

$$= \int_{r^2 \leq t^2} \int_{\theta \in [0, 2\pi]} \frac{1}{2\pi} e^{-r^2/2} r \, dr \, d\theta = \int_0^t e^{-r^2/2} r \, dr \underbrace{\int_0^{2\pi} \frac{1}{2\pi} d\theta}_{=1}$$

substitution $u = r^2/2$, $du = r \, dr$

$$= \int_0^t e^{-u} \, du \cdot 1 = 1 - e^{-t}.$$

Med $t = 2$ fås $P(R \leq 2) = 1 - e^{-2} = 0.8647$.

Notera att R är exponentialfördelad med väntevärde 1 och att $R^2 = X^2 + Y^2$ är $\chi^2(2)$ -fördelad.

6.29 Både X och Y är normalfördelade med väntevärde 180 och varianser 64, dvs standardavvikelse 8.

Den stokastiska variabeln $Y|X = x$, det vill säga den stokastiska variabel som har täthetsfunktion $f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, är normalfördelad med väntevärde $\mu_y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ och varians $\sigma_Y^2(1 - \rho^2)$. Här är det betingade väntevärdet

$$E(Y|X = x) = \mu_y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 180 + \rho \frac{8}{8}(x - 180) = 180(1 - \rho) + \rho x = 90 + 0.5x$$

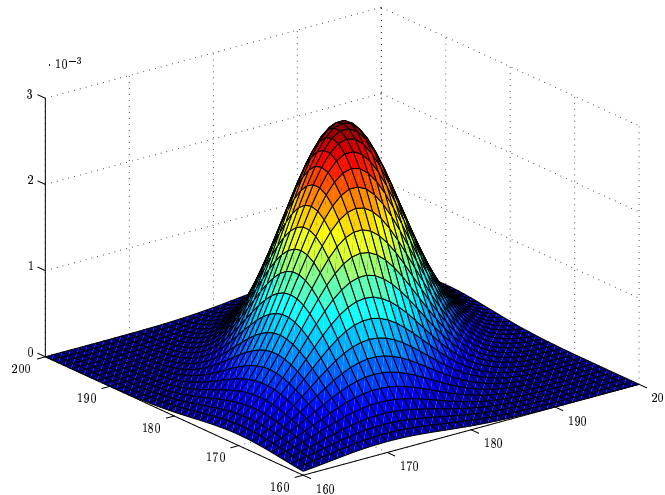
en rät linje genom (180, 180) med lutningen $\rho = 0.5$. Av symmetriskäl är även $E(X|Y = y) = 90 + 0.5y$ samma linje.

Om $X = 190$ är $Y|X = 190$ normalfördelad med väntevärde $90 + 0.5 \cdot 190 = 185$ och varians

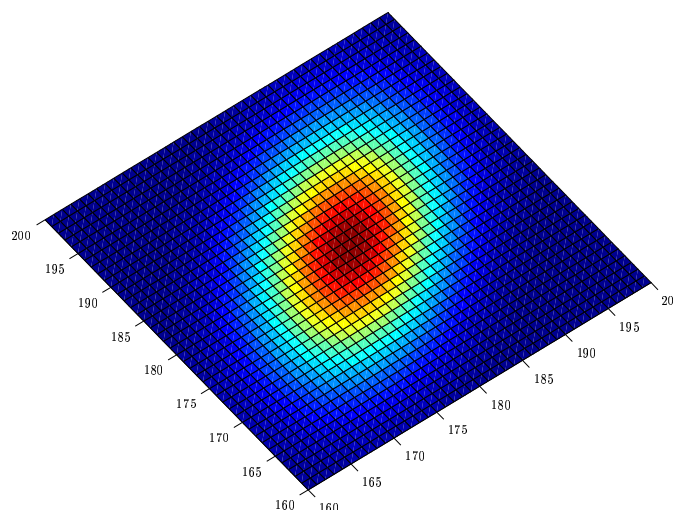
$$\sigma_Y^2(1 - \rho^2) = 64 \cdot (1 - (0.5)^2) = 48$$

det vill säga $Y|X = 190$ är $N(185, \sqrt{48})$. Då är

$$P(Y > 185 | X = 190) = P\left(\frac{Y - 185}{\sqrt{48}} > \frac{185 - 185}{\sqrt{48}} \mid X = 190\right) = 1 - \Phi(0) = 1 - \frac{1}{2} = \frac{1}{2}.$$



Täthetsfunktionen för den tvådimensionella normalfördelning som beskriver söner och fäders längder.



Samma täthetsfunktion men sedd ovanifrån. Här syns korrelationen mellan längderna tydligt.

7.1 Låt X beskriva antalet defekta bland de $n = 15$ utvalda. Modell: enheter är defekta oberoende av varandra med sannolikhet $p = 0.10$ så X är $\text{Bin}(n, p)$, det vill säga

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

för $k = 0, 1, 2, \dots, n$. Då är

$$E(X) = np = 15 \cdot 0.10 = 1.5$$

och

$$V(X) = np(1-p) = 15 \cdot 0.10 \cdot 0.90 = 1.35,$$

dvs $D(X) = \sqrt{V(X)} = 1.1619$. Vidare så är

$$P(X > 3) = P(X \geq 4) = \sum_{k=4}^n \binom{n}{k} p^k (1-p)^{n-k} = 0.055556$$

alternativt

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^3 \binom{n}{k} p^k (1-p)^{n-k} = 1 - 0.94444 = 0.055556.$$

7.2 Låt Y beskriva antalet klave en person får i 2 oberoende slantsinglingar. Då är Y binomialfördelad, Y är $\text{Bin}(2, 1/2)$.

Att en person får samma resultat i båda sina kast har sannolikhet

$$p = P(\{\text{Båda krona}\} \cup \{\text{Båda klave}\}) = P(Y = 0) + P(Y = 2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Personer singlar oberoende av varandra så av $n = 15$ personer beskrivs antalet personer med samma resultat i sina båda kast av den stokastiska variabeln X som är $\text{Bin}(n, p) = \text{Bin}(15, 0.5)$.

7.3 Antalet svarta kort som ett par drar beskrivs av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(2, 1/2)$.

a) Sannolikheten att ett visst par får någon utdelning är

$$P(X > 0) = 1 - P(X = 0) = 1 - \binom{2}{0} \left(\frac{1}{2}\right)^0 \left(1 - \frac{1}{2}\right)^2 = 1 - \frac{1}{4} = \frac{3}{4}.$$

b) Antalet familjer som får utdelning beskrivs av den stokastiska variabeln Y där Y är $\text{Bin}(5, q)$ och $q = P(X > 0) = \frac{3}{4}$.

c)

$$P(Y = 3) = \binom{5}{3} q^3 (1 - q)^{5-3} = 10 \cdot \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^2 = \frac{135}{512} = 0.2637.$$

7.4 Om X är $\text{Bin}(n, p) = \text{Bin}(7, 3/4)$ så är

$$\begin{aligned} P(X \leq 3) &= \sum_{k=0}^3 P(X = k) = \sum_{k=0}^3 \binom{7}{k} p^k (1 - p)^{7-k} = \sum_{k=0}^3 \binom{7}{k} \left(\frac{3}{4}\right)^k \left(1 - \frac{3}{4}\right)^{7-k} \\ &= \frac{1}{16384} + \frac{21}{16384} + \frac{189}{16384} + \frac{945}{16384} = \frac{289}{4096} = 0.0706. \end{aligned}$$

7.5 Om X är $\text{Bin}(n, p) = \text{Bin}(16, 0.40)$ så är

$$\begin{aligned} P(X = 5) &= \binom{16}{5} (0.40)^5 (1 - 0.40)^{16-5} = 0.16227 \\ P(X = 6) &= \binom{16}{6} (0.40)^6 (1 - 0.40)^{16-6} = 0.19833 \\ P(X = 7) &= \binom{16}{7} (0.40)^7 (1 - 0.40)^{16-7} = 0.18889 \end{aligned}$$

så

$$P(4 < X < 8) = P(5 \leq X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7) = 0.54950.$$

7.6 a) Ett försök lyckas med sannolikheten $p = 0.80$. Av $n = 12$ oberoende försök låt X beskriva antalet lyckade. Då är X binomialfördelad, $\text{Bin}(n, p) = \text{Bin}(12, 0.80)$.

b) Ett försök misslyckas med sannolikheten $q = 1 - p = 0.20$. Av $n = 12$ oberoende försök är antalet misslyckade försök $Y = n - X$ binomialfördelad, $\text{Bin}(n, q) = \text{Bin}(12, 0.20)$.

c) Nu är

$$\begin{aligned} P(2 \leq Y \leq 4) &= P(Y = 2) + P(Y = 3) + P(Y = 4) \\ &= \binom{12}{2} q^2 (1 - q)^{12-2} + \binom{12}{3} q^3 (1 - q)^{12-3} + \binom{12}{4} q^4 (1 - q)^{12-4} \\ &= 0.28347 + 0.23622 + 0.13288 = 0.65257. \end{aligned}$$

d) En liknande uträkning som i c) kan användas för att bestämma $P(7 < X \leq 10)$. Vi kan även utnyttja sambandet mellan X och Y , sambandet mellan antalet lyckade och antalet misslyckade försök: $X + Y = n$. Alltså

$$\begin{aligned} P(7 < X \leq 10) &= P(X = 8) + P(X = 9) + P(X = 10) \\ &= P(Y = n - 8) + P(Y = n - 9) + P(Y = n - 10) \\ &= P(Y = 4) + P(Y = 3) + P(Y = 2) = 0.65257. \end{aligned}$$

7.7 Antag att ett frö gror med sannolikhet $p = 0.75$ oberoende av andra frön. Av $n = 15$ sådda frön låt X beskriva antalet frön som gror. Då är X binomialfördelad, X är $\text{Bin}(n, p)$ och

$$P(0.65n \leq X \leq 0.90n) = P(9.75 \leq X \leq 13.5) = \sum_{k=10}^{13} P(X = k) = \sum_{k=10}^{13} \binom{15}{k} p^k (1 - p)^{15-k} = 0.77145.$$

Om man har tillgång till tabell för fördelningsfunktionen för binomialfördelningen kan man utnyttja den för att bestämma sannolikheten ovan.

$$P(0.65n \leq X \leq 0.90n) = P(9 < X \leq 13) = F_X(13) - F_X(9) = 0.91982 - 0.14837 = 0.77145.$$

7.8 Låt X vara antalet \square eller \boxplus i $n = 288$ oberoende tärningskast.

- Då är X binomialfördelad, X är $\text{Bin}(n, p)$ med $n = 288$ och $p = P(\square \text{ eller } \boxplus \text{ i ett tärningskast}) = \frac{1}{3}$.
- $E(X) = np = 288 \cdot \frac{1}{3} = 96 = \mu$ och $V(X) = np(1-p) = 64$ så $D(X) = \sqrt{V(X)} = \sqrt{64} = 8 = \sigma$.
- Enligt tumregeln $V(X) = 64 \geq 10$ kan binomialfördelningen approximeras med normalfördelningen.
- Sannolikheten $P(X \leq 100)$ är med en exakt uträkning

$$\begin{aligned} P(X \leq 100) &= \sum_{k=0}^{100} P(X = k) = \sum_{k=0}^{100} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{100} \binom{288}{k} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{288-k} \\ &= 0.71469. \end{aligned}$$

Med normalapproximation fås

$$P(X \leq 100) = P\left(\frac{X - \mu}{\sigma} \leq \frac{100 - \mu}{\sigma}\right) \approx \Phi\left(\frac{100 - 96}{8}\right) = \Phi(0.5) = 0.6915.$$

Används dessutom halvkorrektion erhålls

$$P(X \leq 100) = P(X \leq 100.5) = P\left(\frac{X - \mu}{\sigma} \leq \frac{100.5 - \mu}{\sigma}\right) \approx \Phi(0.5625) = 0.7131.$$

Sannolikheten $P(X > 80)$ är med en exakt uträkning

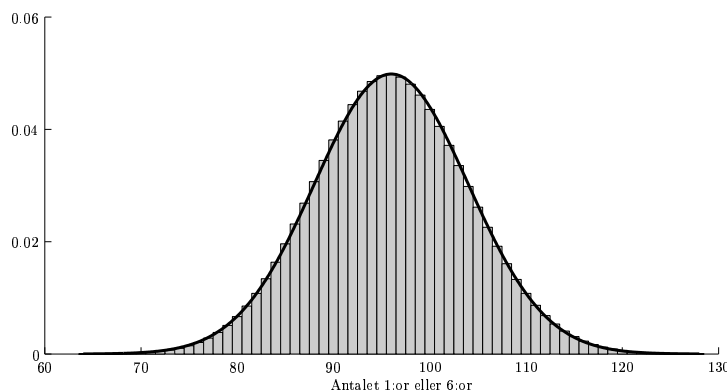
$$\begin{aligned} P(X > 80) &= \sum_{k=81}^{288} P(X = k) = \sum_{k=81}^{288} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=81}^{288} \binom{288}{k} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{288-k} \\ &= 0.97495. \end{aligned}$$

Med normalapproximation fås

$$P(X > 80) = P\left(\frac{X - \mu}{\sigma} > \frac{80 - \mu}{\sigma}\right) \approx 1 - \Phi\left(\frac{80 - 96}{8}\right) = 1 - \Phi(-2) = \Phi(2) = 0.9773.$$

Används dessutom halvkorrektion erhålls

$$\begin{aligned} P(X > 80) &= P(X > 80.5) = P\left(\frac{X - \mu}{\sigma} > \frac{80.5 - \mu}{\sigma}\right) \approx 1 - \Phi(-1.9375) \\ &= \Phi(1.9375) = 0.9737. \end{aligned}$$



Antalet \square eller \boxplus i 288 tärningskast. Binomialfördelningen, ritat som grå staplar, tillsammans med approximerande normalfördelningskurva ritad som heldragen linje.

7.9 (Jämför med uppgift 7.2.)

Låt Y beskriva antalet klave i fem slantsinglingar. Då är Y binomialfördelad, Y är $\text{Bin}(5, 1/2)$. Att en person får samma resultat i alla fem kast har sannolikhet

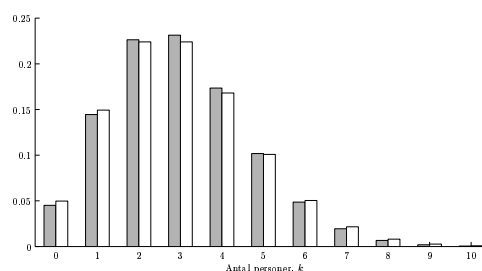
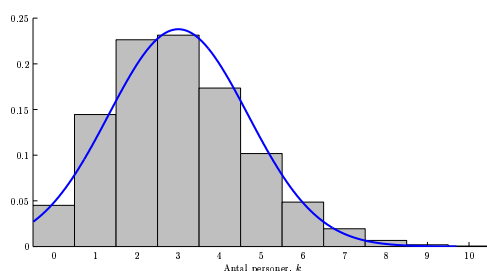
$$p = P(\{\text{Alla krona}\} \cup \{\text{Alla klave}\}) = P(Y = 0) + P(Y = 5) = \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{1}{2^4} = \frac{1}{16}.$$

a) Personer singlar oberoende av varandra så av $n = 48$ personer beskrivs antalet personer med samma resultat i sina båda kast av den stokastiska variabeln X som är $\text{Bin}(n, p) = \text{Bin}(48, 1/16)$.

b) I binomialfördelningen så är

$$E(X) = np = 48 \cdot \frac{1}{16} = 3 \quad V(X) = np(1-p) = 48 \cdot \frac{1}{16} \cdot \frac{15}{16} = \frac{45}{16}.$$

c) Tumregeln för normalapproximation av binomialfördelningen är att $V(X) = np(1-p) > 10$. Här är $V(X) \approx 2.81$ så normalapproximation vore olämpligt. Tumregeln för Poissonapproximation av binomialfördelningen är att $p < 0.10$. Här är $p = 1/16$ så Poissonapproximation är inte olämpligt.



Till vänster är binomialfördelningen inritad med tätheten för en normalfördelning med samma väntevärde och varians. Arean av binomialfördelningens staplar stämmer dåligt överens med motsvarande area under normalfördelningskurvan.

Till höger är binomialfördelningen (grå staplar) inritade tillsammans med Poissonfördelningen (vita staplar). Det är en hyfsad överensstämmelse mellan fördelningarna.

d) För binomialfördelningen har man att

$$P(X \leq 3) = \sum_{k=0}^3 P(X = k) = \sum_{k=0}^3 \binom{n}{k} p^k (1-p)^{n-k} = 0.64731.$$

Med Poissonapproximation, $\mu = np = 3$, erhålls

$$P(X \leq 3) = \sum_{k=0}^3 P(X = k) \approx \sum_{k=0}^3 \frac{\mu^k}{k!} e^{-\mu} = 0.64723.$$

På samma sätt kan sannolikheten

$$P(X > 4) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^4 \binom{n}{k} p^k (1-p)^{n-k} = 1 - 0.82083 = 0.17917$$

approximeras med

$$1 - \sum_{k=0}^4 \frac{\mu^k}{k!} e^{-\mu} = 1 - 0.81526 = 0.18474.$$

- 7.10** Låt X_1 beskriva antalet defekta varvräknare i urvalet från maskin A och X_2 motsvarande för maskin B. Modell: X_1 och X_2 är oberoende och X_i är $\text{Bin}(n_i, p_i)$. Det totala antalet defekta varvräknare i kartongen är $X = X_1 + X_2$ med

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = n_1 p_1 + n_2 p_2$$

och

$$V(X) = V(X_1 + X_2) = \{\text{ober.}\} = V(X_1) + V(X_2) = n_1 p_1 (1 - p_1) + n_2 p_2 (1 - p_2).$$

Notera att X inte är binomialfördelad såvida inte $p_1 = p_2$. Då $p_1 = p_2 = p$ är $X \sim \text{Bin}(n_1 + n_2, p)$.

- 7.11** Låt q vara sannolikheten att en flygplansmotor fungerar och antag att flygplansmotorer fungerar oberoende av varandra. Med X som antalet fungerande motorer på ett tvåmotorigt plan så är X binomialfördelad med $n = 2$ och

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - q)^2.$$

Med Y som antalet fungerande motorer på ett fyrmotorigt plan så är Y binomialfördelad med $n = 4$ och

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - (1 - q)^4 - 4q(1 - q)^3.$$

Olikheten $P(X \geq 1) \geq P(Y \geq 2)$ ger att

$$\begin{aligned} 1 - (1 - q)^2 &\geq 1 - (1 - q)^4 - 4q(1 - q)^3 \\ (1 - q)^4 + 4q(1 - q)^3 &\geq (1 - q)^2 \\ (1 - q)^2 + 4q(1 - q) &\geq 1 \\ 1 - 2q + q^2 + 4q - 4q^2 &\geq 1 \\ 2 - 3q &\geq 0 \\ 2/3 &\geq q \end{aligned}$$

Med $p = 1 - q$ som sannolikheten för att en motor inte fungerar blir kravet $2/3 \geq q$ att $p \geq 1/3$.

- 7.12** Låt X beskriva antalet tentander av n tentarnde som klarar tentamen. Då är X binomialfördelad, X är $\text{Bin}(n, p)$ med $p = 0.70$.

a) Med $n = 4$ tentander är

$$\begin{aligned} P(X \geq 0.75n) &= P(X \geq 3) = P(X = 3) + P(X = 4) = \binom{4}{3} p^3 (1 - p)^1 + \binom{4}{4} p^4 (1 - p)^0 \\ &= 0.4116 + 0.2401 = 0.6517. \end{aligned}$$

b) Med $n = 16$ tentander är

$$P(X \geq 0.75n) = P(X \geq 12) = \sum_{k=12}^n \binom{n}{k} p^k (1 - p)^{n-k} = 0.4499$$

c) Med $n = 400$ tentander är

$$P(X \geq 0.75n) = P(X \geq 300) = \sum_{k=300}^n \binom{n}{k} p^k (1 - p)^{n-k} = 0.01553.$$

Det sista svaret kan beräknas med normalapproximation eftersom $\sigma^2 = V(X) = np(1 - p) = 84 \geq 10$. Då är $\mu = E(X) = np = 280$ och

$$P(X \geq 300) = P\left(\frac{X - \mu}{\sigma} \geq \frac{300 - \mu}{\sigma}\right) \approx 1 - \Phi\left(\frac{300 - 280}{\sqrt{84}}\right) = 1 - \Phi(2.18) = 0.0146.$$

Med halvkorrektion fås svaret

$$P(X \geq 300) = P(X \geq 299.5) = P\left(\frac{X - \mu}{\sigma} \geq \frac{299.5 - \mu}{\sigma}\right) \approx 1 - \Phi(2.13) = 0.0166.$$

7.13 Låt X beskriva antalet defekta byggelement bland $n = 1000$ stycken. Modell: Ett byggelement är defekt med sannolikhet $p = 0.10$ och oberoende av andra byggelement. X är $\text{Bin}(n, p)$. Då är

$$P(80 \leq X \leq 120) = \sum_{k=80}^{120} P(X = k) = \sum_{k=80}^{120} \binom{n}{k} p^k (1-p)^{n-k} = 0.96948.$$

Med approximation. Vi vet att

$$E(X) = np = 100 = \mu \quad V(X) = np(1-p) = 90 = \sigma^2$$

Då $V(X) \geq 10$ är approximation av binomialfördelningen med normalfördelningen tillåten, dvs X är approximativt $N(\mu, \sigma) = N(100, \sqrt{90})$. Då är

$$\begin{aligned} P(80 \leq X \leq 120) &= P\left(\frac{80 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{120 - \mu}{\sigma}\right) \approx \Phi\left(\frac{120 - \mu}{\sigma}\right) - \Phi\left(\frac{80 - \mu}{\sigma}\right) \\ &= 2\Phi\left(20/\sqrt{90}\right) - 1 = 0.96499. \end{aligned}$$

Med halvkorrektion får man svaret

$$\begin{aligned} P(80 \leq X \leq 120) &= P(79.5 \leq X \leq 120.5) = P\left(\frac{79.5 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{120.5 - \mu}{\sigma}\right) \\ &\approx \Phi\left(\frac{120.5 - \mu}{\sigma}\right) - \Phi\left(\frac{79.5 - \mu}{\sigma}\right) = 2\Phi\left(20.5/\sqrt{90}\right) - 1 = 0.96930. \end{aligned}$$

7.14 Låt X_A beskriva antalet användbara enheter i urvalet från maskin A och X_B motsvarande för maskin B. Eftersom enheter är användbara oberoende av tidigare producerade enheter ansätter vi modellen att

$$X_A \text{ är } \text{Bin}(n_A, p_A) \quad \text{och} \quad X_B \text{ är } \text{Bin}(n_B, p_B)$$

där X_A och X_B är oberoende och $p_A = 1 - 0.05 = 0.95$ och $p_B = 1 - 0.10 = 0.90$.

Med $n_A = 1200$ och $n_B = 1300$ så är $V(X_A) = n_A p_A (1 - p_A) = 57$ och $V(X_B) = n_B p_B (1 - p_B) = 117$ vilka båda är större än 10 vilket medger normalapproximation av binomialfördelningarna.

Vi söker $P(X_A > X_B) = P(X_A - X_B > 0)$ där $X_A - X_B$ är approximativt normalfördelad med väntevärde

$$\mu = E(X_A - X_B) = E(X_A) - E(X_B) = -30$$

och varians

$$\sigma^2 = V(X_A - X_B) = \{\text{oberoende}\} = V(X_A) + (-1)^2 V(X_B) = 57 + 117 = 174,$$

det vill säga $\sigma = 13.191$. Alltså är

$$\begin{aligned} P(X_A > X_B) &= P(X_A - X_B > 0) = P\left(\frac{(X_A - X_B) - \mu}{\sigma} > \frac{0 - \mu}{\sigma}\right) \approx 1 - \Phi\left(-\frac{\mu}{\sigma}\right) \\ &= 1 - \Phi(2.2743) = 1 - 0.98853 = 0.01147. \end{aligned}$$

7.15 Antag att en tillverkad enheter är defekt med sannolikhet $p = 0.005$ oberoende av andra tillverkade enheter. Av $n = 100$ tillverkade enheter låt X beskriva antalet defekta. Då är X binomialfördelad, X är $\text{Bin}(n, p)$ och

$$q = P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^3 P(X = k) = 1 - \sum_{k=0}^3 \binom{n}{k} p^k (1-p)^{n-k} = 1 - 0.99833 = 0.00167.$$

Eftersom $p < 0.10$ är en Poissonapproximation av binomialfördelningen tillåten och approximativt erhålles med $\mu = np = 0.5$

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^3 P(X = k) \approx 1 - \sum_{k=0}^3 \frac{\mu^k}{k!} e^{-\mu} = 1 - 0.99825 = 0.00175.$$

Alltså är en kartong med 100 enheter dålig med sannolikhet q oberoende av andra kartonger. Av $m = 10\,000$ kartonger är antalet dåliga kartonger, Y , binomialfördelat, det vill säga Y är $\text{Bin}(m, q)$. Alltså är

$$P(Y > 25) = 1 - P(Y \leq 25) = 1 - \sum_{k=0}^{25} P(Y = k) = 1 - \sum_{k=0}^{25} \binom{m}{k} q^k (1-q)^{m-k} = 1 - 0.97871 = 0.02129.$$

Eftersom $q < 0.10$ är en Poissonapproximation av binomialfördelningen tillåten och approximativt erhålles med $\mu_2 = mq = 16.733$

$$P(Y > 25) = 1 - P(Y \leq 25) = 1 - \sum_{k=0}^{25} P(Y = k) \approx 1 - \sum_{k=0}^{25} \frac{\mu_2^k}{k!} e^{-\mu_2} = 1 - 0.97862 = 0.02138.$$

Än enklare är förmodligen att utnyttja normalapproximation av binomialfördelningen. Här är $V(Y) = mq(1-q) \approx 16.7 > 10$ så normalapproximation kan göras. Med $\mu_2 = mq$ och $\sigma = \sqrt{mq(1-q)}$ erhålles

$$P(Y > 25) = P\left(\frac{Y - \mu_2}{\sigma} > \frac{25 - \mu_2}{\sigma}\right) \approx 1 - \Phi\left(\frac{25 - \mu_2}{\sigma}\right) = 1 - \Phi(2.0228) = 1 - 0.97845 = 0.02155.$$

7.16 (Jämför med uppgift 2.21). Av $N = 100$ distinkta enheter är $s = 6$ defekta, det vill säga andelen defekta enheter är $p = 6\%$. Av $n = 5$ på måfå utvalda enheter låt X vara antalet defekta om urvalet skedde utan återläggning. Då är X hypergeometriskt fördelat och

$$P(X = k) = \frac{\binom{\text{\#sätt att välja } k \text{ bland de defekta}}{\binom{\text{\#sätt att välja } 5 - k \text{ bland de hela}}{\binom{6}{k} \binom{94}{5 - k}}}{\text{\#sätt att välja 5 bland alla}} = \frac{\binom{6}{k} \binom{94}{5 - k}}{\binom{100}{5}}.$$

för $k = 0, 1, \dots, 5$. Med andelen p som parameter är

$$E(X) = np = 5 \cdot 0.06 = 0.30 \quad V(X) = np(1-p) \frac{N-n}{N-1} = 5 \cdot 0.06 \cdot 0.94 \frac{100-5}{100-1} = 0.27061$$

dvs $D(X) = \sqrt{V(X)} = 0.5202$. Vidare så är

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{54891018 + 18297006}{75287520} = \frac{435643}{448140} = 0.97211.$$

7.17 I en population av storlek N är andelen p män, det vill säga antalet män i populationen är Np . Om n personer väljes på måfå är antalet utvalda män, X ,

- hypergeometriskt fördelat om urvalet sker utan återläggning. Då är

$$E(X) = np \quad V(X) = np(1-p) \frac{N-n}{N-1}.$$

- binomialfördelat om urvalet sker med återläggning. Då är

$$E(X) = np \quad V(X) = np(1-p).$$

Med värden $n = 100$, $p = 0.60$ erhålles

	$E(X)$	$V(X)$	
		med	utan
$N = 200$	60	24	12.06
$N = 1000$	60	24	21.62

7.18 a) För att hypergeometrisk fördelning skall kunna approximeras med binomialfördelning skall populationens storlek N vara mycket större än urvalsstorleken n . Vi kräver att $n/N < 0.10$. Här är $n/N = 0.01$ så approximationen är tillåten.

b) För att hypergeometrisk fördelning skall kunna approximeras med normalfördelning skall variansen inte vara för liten. Vi kräver att $V(X) = np(1-p)d_n^2 > 10$. Här är variansen 78 så approximationen är tillåten.

c) Tvåstegsapproximation. Eftersom $n/N = 0.05 < 0.10$ så kan fördelningen approximeras med binomialfördelningen. Denna binomialfördelning låter sig approximeras med normalfördelningen om dess varians $np(1-p) > 10$. Här är $np(1-p) = 345$ så normalapproximation är tillåten. (Bättre är dock att göra approximationen i ett steg och använda $np(1-p)d_n^2$ som varians i normalfördelningen.

d) Enstegsapproximation. Approximationen är tillåten om $p + n/N < 0.10$. Här är $p + n/N = 0.065$ så approximationen är tillåten.

Tvåstegsapproximation. Eftersom $n/N = 0.025 < 0.10$ så kan fördelningen approximeras med binomialfördelningen. Denna binomialfördelning låter sig approximeras med Poissonfördelningen om $p < 0.10$. Här är $p = 0.04$ så Poissonapproximation är tillåten.

7.19 Låt p vara andelen defekta i partiet om storlek $N = 1000$ varor. Då är $p = s/N$ för något heltal s , antalet defekta varor. Om $n = 30$ varor väljes på måfå utan återläggning och X beskriver antalet defekta bland dessa så är X hypergeometriskt fördelat med

$$P(X = k) = \frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n}} = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}},$$

där $0 \leq k \leq s$ och $0 \leq n-k \leq N-s$. Eftersom $n/N < 0.10$ så kan fördelningen för X approximeras med binomialfördelningen, dvs

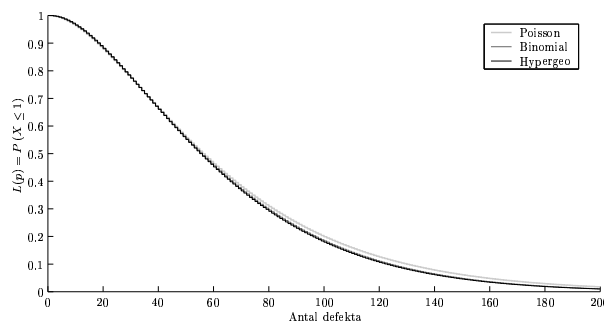
$$P(X = k) \approx \binom{n}{k} p^k (1-p)^{n-k}$$

för $k = 0, \dots, n$, och då $p < 0.10$ kan binomialfördelningen approximeras med Poissonfördelningen med väntevärde np så

$$P(X = k) \approx \frac{(np)^k}{k!} e^{-np}$$

för $k = 0, 1, \dots$. Nu är

$$L(p) = P(X \leq 1) = P(X = 0) + P(X = 1) \approx \frac{(np)^0}{0!} e^{-np} + \frac{(np)^1}{1!} e^{-np} = (1 + np)e^{-np}.$$



7.20 Av $N = 1000$ radioapparater är $s = 12$ defekta, det vill säga andelen defekta enheter är $p = 1.2\%$. Av $n = 50$ på måfå utvalda enheter låt X vara antalet defekta om urvalet skedde utan återläggning. Då är X hypergeometriskt fördelad och

$$P(X = k) = \frac{\binom{12}{k} \binom{1000 - 12}{50 - k}}{\binom{1000}{50}}$$

för $k = 0, 1, \dots, 12$.

Då är

$$\begin{aligned} P(\text{Partiet accepteras}) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.69242 + 0.25993 + 0.043186 = 0.99553. \end{aligned}$$

Eftersom $n/N = 50/1000 = 0.05 < 0.10$ så kan den hypergeometriskta fördelningen approximeras med en binomialfördelning. Med binomialapproximation fås svaret

$$\begin{aligned} P(\text{Partiet accepteras}) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{30}{0} p^0 (1-p)^{30} + \binom{30}{1} p^1 (1-p)^{29} + \binom{30}{2} p^2 (1-p)^{28} \\ &= 0.69616 + 0.25366 + 0.044673 = 0.99449. \end{aligned}$$

Eftersom $p < 0.10$ kan binomialfördelningen approximeras med en Poissonfördelning. Gör denna approximation fås, med $\mu = np = 0.36$, svaret

$$\begin{aligned} P(\text{Partiet accepteras}) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{\mu^0}{0!} e^{-\mu} + \frac{\mu^1}{1!} e^{-\mu} + \frac{\mu^2}{2!} e^{-\mu} = 0.99405. \end{aligned}$$

7.21 Om X är $\text{Po}(\mu)$ med $\mu = 7.5$ så är

$$P(X \leq 4) = \sum_{k=0}^4 P(X = k) = \sum_{k=0}^4 \frac{\mu^k}{k!} e^{-\mu} = \sum_{k=0}^4 \frac{\mu^k}{k!} e^{-\mu} = 0.1321.$$

Vidare så är

$$P(6 \leq X \leq 11) = \sum_{k=6}^{11} P(X = k) = \sum_{k=6}^{11} \frac{\mu^k}{k!} e^{-\mu} = 0.6793,$$

$$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{k=0}^9 P(X = k) = 1 - \sum_{k=0}^9 \frac{\mu^k}{k!} e^{-\mu} = 1 - 0.7764 = 0.2236$$

och

$$P(X = 8) = \frac{\mu^8}{8!} e^{-\mu} = 0.1373.$$

7.22 För en Poissonfördelad stokastisk variabel X med parameter μ är $E(X) = \mu$ och $V(X) = \mu$. Variationskoefficienten

$$\frac{D(X)}{E(X)} = \frac{\sqrt{\mu}}{\mu} = \frac{1}{\sqrt{\mu}} = \frac{1}{2}$$

vilket ger $\mu = 4$. Då är

$$P(X = 0) = \frac{\mu^0}{0!} e^{-\mu} = e^{-\mu} = 0.018316.$$

7.23 Låt Y vara antalet registreringar under ett tidsintervall av längd t sekunder. Med modellen att Y är Poissonfördelad med väntevärde λt så är $P(Y > 1.10\lambda t) = \dots$

- a) $\dots = P(Y > 8.8) = 1 - P(Y \leq 8) = 1 - 0.5925 = 0.4075$
 b) $\dots = P(Y > 15.4) = 1 - P(Y \leq 15) = 1 - 0.6694 = 0.3306$
 c) $\dots = P(Y > 110) = 1 - P(Y \leq 110) = 1 - 0.8529 = 0.1471$

7.24 Låt X_1, \dots, X_4 beskriva antalet samtal till telefon 1, $\dots, 4$ under tidsintervallet. Modell: X_i är oberoende och Poissonfördelade med väntevärden

$$E(X_1) = E(X_2) = E(X_3) = 1 \quad E(X_4) = 0.5.$$

Med Y som det totala antalet samtal till kontoret är

$$Y = X_1 + X_2 + X_3 + X_4$$

en Poissonfördelad stokastisk variabel med väntevärde

$$m = E(Y) = E(X_1 + X_2 + X_3 + X_4) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 3.5.$$

Sökt är

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y < 3) = 1 - P(Y \leq 2) = 1 - \sum_{k=0}^2 P(Y = k) = 1 - \sum_{k=0}^2 \frac{m^k}{k!} e^{-m} \\ &= 1 - 0.32085 = 0.67915. \end{aligned}$$

7.25 Låt X beskriva antalet spår i emulsionen och ansätt modellen att X är en $Po(\mu)$ -fördelad stokastisk variabel.

Om $0.07 = P(X = 0) = e^{-\mu}$ så är $\mu = -\ln(0.07) = 2.6593$. Sannolikhetsfunktionen $p_X(k)$, $k = 0, 1, 2, \dots$, med detta parametervärde

$k :$	0	1	2	3	4	5	\dots
$p_X(k) :$	0.0700	0.1861	0.2475	0.2194	0.1459	0.0776	\dots

maximeras för $k = 2$. Typvärdet i fördelningen, det i modellen vanligaste förekommande värdet, är således 2.

7.26 Låt X beskriva antalet blodkroppar i 1mm^3 blod hos personen. Då är X Poissonfördelad med parameter $\lambda = 6000$. Då är

$$P(X < 5000) = \sum_{k=0}^{4999} P(X = k) = \sum_{k=0}^{4999} \frac{\lambda^k}{k!} e^{-\lambda} = 1.1 \cdot 10^{-40}.$$

Eftersom $\lambda > 15$ är normalapproximation av Poissonfördelningen tillåten. Med normalapproximation fås,

$$P(X < 5000) = P\left(\frac{X - \lambda}{\sqrt{\lambda}} < \frac{5000 - \lambda}{\sqrt{\lambda}}\right) = \Phi\left(\frac{5000 - \lambda}{\sqrt{\lambda}}\right) = \Phi(-12.91) = 2.0 \cdot 10^{-38}.$$

Antalet blodkroppar i en $\text{ml} = 1000 \text{mm}^3$ blod är Poissonfördelat med väntevärde 1000λ . Dessa blandas upp i en vätska om 1 liter = 10^6mm^3 . Om en mm^3 vätska plockas ut ansätter vi modellen att varje blodkropp i vätskan har sannolikheten $p = 1/10^6$ tas med i urvalet. Antalet blodkroppar i urvalet, Y , blir då Poissonfördelat med parameter $\mu = (1000\lambda)p = 6$. Då är

$$P(Y < 5) = \sum_{k=0}^4 P(Y = k) = \sum_{k=0}^4 \frac{\mu^k}{k!} e^{-\mu} = 0.28506.$$

7.27 Låt X_1, \dots, X_n beskriva antalet larm under dagar $1, \dots, n$. Modell: X_i är oberoende och Poissonfördelade. Det totala antalet larm under ett år är

$$Y = \sum_{i=1}^n X_i,$$

en summa av oberoende Poissonfördelade stokastiska variabler, alltså är Y Poissonfördelad med parameter

$$\mu = E(Y) = E\left(\sum_{i=1}^n X_i\right) = nE(X_i) = 365 \cdot \frac{1}{2} = 182.5.$$

Nu är

$$P(Y > 200) = 1 - P(Y \leq 200) = 1 - \sum_{k=0}^{200} P(Y = k) = 1 - \sum_{k=0}^{200} \frac{\mu^k}{k!} e^{-\mu} = 1 - 0.90717 = 0.09283.$$

Eftersom $\mu \geq 15$ så kan fördelningen för Y approximeras med en normalfördelning, $N(\mu, \sqrt{\mu})$. Då är

$$P(Y > 200) = P\left(\frac{Y - \mu}{\sqrt{\mu}} > \frac{200 - \mu}{\sqrt{\mu}}\right) \approx 1 - \Phi\left(\frac{200 - \mu}{\sqrt{\mu}}\right) = 1 - \Phi(1.2954) = 1 - 0.90241 = 0.09759.$$

Än bättre blir approximationen om man utnyttjar halvkorrektion. Då är

$$\begin{aligned} P(Y > 200) &= P(Y > 200.5) = P\left(\frac{Y - \mu}{\sqrt{\mu}} > \frac{200.5 - \mu}{\sqrt{\mu}}\right) \approx 1 - \Phi\left(\frac{200.5 - \mu}{\sqrt{\mu}}\right) = 1 - \Phi(1.3324) \\ &= 1 - 0.90864 = 0.09136. \end{aligned}$$

7.28 Låt X beskriva antalet flygplan av $n = 100$ som totalhavererar under ett år. Modell: plan totalhavererar med sannolikhet $p = 0.008$ och oberoende av varandra. X är $\text{Bin}(n, p)$. Försäkringsbolaget vinst ges av

$$Y = n \cdot 10^4 - X \cdot 10^6 = 10^6(1 - X).$$

Händelsen $Y < 0$ är samma händelse som $X > 1$ och

$$\begin{aligned} P(Y < 0) &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \left(\binom{n}{0} p^0 (1-p)^{100} + \binom{n}{1} p^1 (1-p)^{99} \right) = 1 - 0.80908 = 0.19092. \end{aligned}$$

Poissonapproximation av Binomialfördelningen går bra om $p < 0.10$, så här kan X sägas vara approximativt Poissonfördelad med parameter $\mu = E(X) = np = 0.8$. Då är

$$1 - (P(X = 0) + P(X = 1)) \approx 1 - \left(\frac{\mu^0}{0!} e^{-\mu} + \frac{\mu^1}{1!} e^{-\mu} \right) = 1 - 0.80879 = 0.19121.$$

7.29 Låt X vara antalet krona en person får vid två singlar. Då är $X \text{ Bin}(n, p) = \text{Bin}(2, 0.5)$, det vill säga med $p_i = P(X = i)$ fås

$$\begin{aligned} p_0 &= P(X = 0) = \binom{2}{0} p^0 (1-p)^2 = \frac{1}{4} \\ p_1 &= P(X = 1) = \binom{2}{1} p^1 (1-p)^1 = \frac{1}{2} \\ p_2 &= P(X = 2) = \binom{2}{2} p^2 (1-p)^0 = \frac{1}{4} \end{aligned}$$

Av $r = 9$ personer som singlar två mynt, låt Y_i vara antalet personer som får precis i krona. Då är $Y_i \text{ Bin}(r, p_i)$ och vektorn (Y_0, Y_1, Y_2) är Multinomialfördelad med parametrar $(r, (p_0, p_1, p_2))$. Vi har att

$$P(Y_0 = 3, Y_1 = 4, Y_2 = 2) = \binom{r}{3, 4, 2} p_0^3 p_1^4 p_2^2 = \frac{9!}{3!4!2!} \left(\frac{1}{4}\right)^3 \left(\frac{1}{2}\right)^4 \left(\frac{1}{4}\right)^2 = \frac{315}{4096} = 0.0769.$$

7.30 Om vektorn (X_1, \dots, X_r) är multinomialfördelad med parametrar $(n, (p_1, \dots, p_r))$ så är komponenterna binomialfördelade. Vidare så är

$$\begin{aligned} Y_1 &= X_1 + X_4 & \text{Bin}(n, p_1 + p_4) &= \text{Bin}(n, 1/2) \\ Y_2 &= X_3 & \text{Bin}(n, p_3) &= \text{Bin}(n, 1/6) \\ Y_3 &= X_2 + X_5 & \text{Bin}(n, p_2 + p_5) &= \text{Bin}(n, 1/3) \end{aligned}$$

och vektorn (Y_1, Y_2, Y_3) är multinomialfördelad med parametrar $(n, (q_1, q_2, q_3)) = (n, (1/2, 1/6, 1/3))$, det vill säga

$$P(Y_1 = k_1, Y_2 = k_2, Y_3 = k_3) = \binom{n}{k_1, k_2, k_3} q_1^{k_1} q_2^{k_2} q_3^{k_3} = \frac{n!}{k_1! k_2! k_3!} \left(\frac{1}{2}\right)^{k_1} \left(\frac{1}{6}\right)^{k_2} \left(\frac{1}{3}\right)^{k_3}$$

där $k_i \geq 0$ och $k_1 + k_2 + k_3 = n$.

7.31 Låt X vara Poissonfördelad med parameter λ och beskriva antalet ägg som insekten lägger. Varje ägg kläcks med sannolikhet p oberoende av andra ägg. Låt Y vara antalet ägg som kläcks. Då är $\Omega_Y = \{0, 1, 2, 3, \dots\}$ och givet att $X = n$ är antalet ägg som kläcks binomialfördelat, $\text{Bin}(n, p)$. Alltså, för $k = 0, 1, 2, \dots$ är

$$\begin{aligned} P(Y = k) &= \sum_{n=k}^{\infty} P(Y = k | X = n) P(X = n) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \frac{(p\lambda)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} = \frac{(p\lambda)^k}{k!} e^{-\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{(\lambda(1-p))^n}{n!}}_{=e^{\lambda(1-p)}} = \frac{(p\lambda)^k}{k!} e^{-\lambda} e^{\lambda(1-p)} \\ &= \frac{(p\lambda)^k}{k!} e^{-p\lambda} = \frac{\mu^k}{k!} e^{-\mu}, \end{aligned}$$

det vill säga Y är Poissonfördelad med parameter $\mu = p\lambda$.

8.1 Med $x_1 = 6$ är

$$\begin{aligned} x_2 &= (17 \cdot x_1) \bmod(7) = 102 \bmod(7) = 4 \\ x_3 &= (17 \cdot x_2) \bmod(7) = 68 \bmod(7) = 5 \\ x_4 &= (17 \cdot x_3) \bmod(7) = 85 \bmod(7) = 1 \\ x_5 &= (17 \cdot x_4) \bmod(7) = 17 \bmod(7) = 3 \\ x_6 &= (17 \cdot x_5) \bmod(7) = 51 \bmod(7) = 2 \\ x_7 &= (17 \cdot x_6) \bmod(7) = 34 \bmod(7) = 6 = x_1 \end{aligned}$$

8.2 Fördelningsfunktionen för X kan skrivas

$$F_X(t) = P(X \leq t) = \begin{cases} 0 & \text{om } t < 1 \\ 0.1 & \text{om } 1 \leq t < 2 \\ 0.4 & \text{om } 2 \leq t < 3 \\ 0.6 & \text{om } 3 \leq t < 4 \\ 1 & \text{om } t \geq 4. \end{cases}$$

Inversen $F_X^{-1}(u) = \min\{t : F(x) \geq u\}$ blir således

$$F_X^{-1}(u) = \begin{cases} 1 & \text{om } 0 < u \leq 0.1 \\ 2 & \text{om } 0.1 < u \leq 0.4 \\ 3 & \text{om } 0.4 < u \leq 0.6 \\ 4 & \text{om } 0.6 < u \leq 1. \end{cases}$$

Slumptalen $u_1 = 0.31$, $u_2 = 0.89$, $u_3 = 0.57$ och $u_4 = 0.27$ genererar utfallen $x_i = F_X^{-1}(u_i)$:

$$\begin{aligned}x_1 &= F_X^{-1}(0.31) = 2 \\x_2 &= F_X^{-1}(0.89) = 4 \\x_3 &= F_X^{-1}(0.57) = 3 \\x_4 &= F_X^{-1}(0.27) = 2.\end{aligned}$$

8.3 Om X är geometriskt fördelad med parameter p är $X + 1$ ffg(p)-fördelad. Alltså kan man erhålla den geometriska fördelningen om man byter ut $y=1$ mot $y=0$ i koden. Nedan följer den modifierade koden:

```
function y=randgeo(p)
% y=randgeo(p) ger ett geometriskt fördelat slumpstal

y=0;
x = rand(1);
while (x>p)
    x=rand(1);
    y=y+1;
end
```

8.4 En stokastisk variabel med täthetsfunktion $f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$ har fördelningsfunktion

$$\begin{aligned}F_X(t) = P(X \leq t) &= \int_{-\infty}^t f_X(x) dx = \int_{-\infty}^t \frac{1}{\pi} \cdot \frac{1}{1+x^2} dx = \frac{1}{\pi} [\tan^{-1}(x)]_{-\infty}^t \\&= \frac{1}{\pi} \left[\tan^{-1}(t) + \frac{\pi}{2} \right] = \frac{\tan^{-1}(t)}{\pi} + \frac{1}{2}.\end{aligned}$$

Detta är en kontinuerlig funktion med inversen

$$F_X^{-1}(u) = \tan \left(\pi \left(u - \frac{1}{2} \right) \right).$$

För att erhålla slumpstal från fördelningen väljs ett slumpstal u , likformigt på intervallet $(0, 1)$ och låter $x = F_X^{-1}(u)$.

Anmärkning: Fördelningen kallas för Cauchy-fördelningen och uppstår bland annat som fördelningen för Z_1/Z_2 där Z_1 och Z_2 är oberoende och $N(0, 1)$. Vi kan således även konstruera utfall från Cauchy-fördelningen med hjälp av två oberoende realiseringar av $N(0, 1)$ -variabler.

8.5 Integralen kan skrivas

$$\begin{aligned}\int_{-\infty}^{\infty} \int_0^{\infty} \cos(xy) e^{-x-y^2} dx dy &= \int_{-\infty}^{\infty} \int_0^{\infty} \cos(xy) e^{-x} e^{-y^2} dx dy = \left\{ \text{med } \sigma = \frac{1}{\sqrt{2}} \right\} \\&= \sqrt{2\pi}\sigma \int_{-\infty}^{\infty} \int_0^{\infty} \cos(xy) e^{-x} \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} dx dy \\&= \sqrt{2\pi}\sigma \int_{-\infty}^{\infty} \int_0^{\infty} \cos(xy) f_X(x) f_Y(y) dx dy \\&= \sqrt{2\pi}\sigma E(\cos(XY))\end{aligned}$$

där X och Y är två oberoende stokastiska variabler sådana att X är exponentialfördelad med väntevärde 1 och Y är normalfördelad med väntevärde 0 och standardavvikelse $\sigma = 1/\sqrt{2}$.

Koden

```
>> n=10000;
>> u=rand(n,1);
>> x=-log(u);
>> z=randn(n,1);
>> y=z/sqrt(2);
```

skapar $n = 10000$ slumpstal x_1, \dots, x_n från $\exp(1)$ -fördelningen av motsvarande antal slumpstal u_1, \dots, u_n likformiga på intervallet $(0, 1)$ samt n slumpstal y_1, \dots, y_n från $N(0, 1/\sqrt{2})$ från n stycken $N(0, 1)$ -slumpstal z_1, \dots, z_n .

Storheten $\sqrt{2\pi}\sigma E(\cos(XY)) = \sqrt{\pi}E(\cos(XY))$ skattas här till

```
>> sqrt(pi)*mean(cos(x.*y))
ans =
    1.3417
```

8.6 Fördelningsfunktionen $F(x) = 1 - \frac{1}{x^a}$ för $x \geq 1$ har invers

$$F^{-1}(u) = (1 - u)^{-1/a}$$

för $0 < u < 1$.

Koden

```
>> a = 3;
>> n = 10000;
>> u1 = rand(n,1);
>> u2 = rand(n,1);
>> x = (1-u1).^(-1/a);
>> y = (1-u2).^(-1/a);
>> z = sum(2*x <= y.^2)
z =
    2340
```

gav en skattning av p som $p_{\text{obs}}^* = 2340/10000 = 0.2340$.

De Paretofördelade stokastiska variablerna har täthetsfunktioner

$$f_X(x) = \frac{d}{dx} F(x) = \frac{a}{x^{a+1}}, \quad x \geq 1.$$

Exakt räkning ger sedan

$$\begin{aligned} p &= P(2X \leq Y^2) = \int \int_{(x,y): 2x \leq y^2} f_{X,Y}(x,y) dy dx = \int \int_{(x,y): 2x \leq y^2} f_X(x) f_Y(y) dy dx \\ &= \int_1^\infty \int_{\sqrt{2x}}^\infty f_X(x) f_Y(y) dy dx = \int_1^\infty \int_{\sqrt{2x}}^\infty \frac{a}{x^{a+1}} \cdot \frac{a}{y^{a+1}} dy dx = \int_1^\infty \frac{a}{x^{a+1}} \left[-\frac{1}{y^a} \right]_{\sqrt{2x}}^\infty dx \\ &= \int_1^\infty \frac{a}{x^{a+1}} \cdot \frac{1}{(2x)^{a/2}} dx = \frac{a}{2^{a/2}} \int_1^\infty \frac{1}{x^{\frac{3}{2}a+1}} dx = \frac{a}{2^{a/2}} \left[-\frac{1}{x^{\frac{3}{2}a}} \cdot \frac{2}{3a} \right]_1^\infty = \frac{2}{3 \cdot 2^{a/2}} = \{a = 3\} \\ &= \frac{1}{3\sqrt{2}} = 0.2357 \end{aligned}$$

Antalet par, z , sådana att $2x \leq y^2$ är ett utfall av en stokastisk variabel Z där Z är $\text{Bin}(n, p)$ med $n = 10000$ och $p = \frac{1}{3\sqrt{2}}$.

8.7 Fördelningsfunktionen $F(t) = e^{-(t/a)^b}$, $t > 0$, har invers

$$F^{-1}(u) = a(-\ln(u))^{-1/b}, \quad 0 < u < 1.$$

Koden

```
>> a = 1;
>> b = 10.4;
>> n = 20000;
>> u = rand(n,1);
>> x = a*(-log(u)).^(-1/b);
```

ger $n = 20000$ utfall från Frechét-fördelningen med de givna parametrarna. Väntevärdet skattas med `mean(x)` till 1.0653 och variansen skattas med `std(x)^2` till 0.0201.

8.8 Koden

```
>> n = 10000;
>> z1 = randn(n,1);
>> z2 = randn(n,1);
```

ger $n = 10000$ par av utfall av oberoende $N(0, 1)$ -fördelade stokastiska variabler. Dessa par används för att skapa observationer (x, y) från multivariat normalfördelade (X, Y) genom att låta

```
>> mux = 0;
>> muy = 0;
>> sigmax = 2;
>> sigmay = 2;
>> rho = 0.5;
>>
>> x = sigmax*z1 + mux;
>> y = sigmay*(rho*z1 + sqrt(1-rho^2)*z2)+muy;
```

Väntevärdet $E\left(\frac{Y^4}{1+X^2}\right)$ skattas till

```
>> mean(y.^4./(1+x.^2))
ans =
    15.3732
```

8.9 Fördelningsfunktionen

$$F(t) = \begin{cases} 0 & \text{om } t < 0 \\ 0.5 + t/16 & \text{om } 0 \leq t < 8 \\ 1 & \text{om } t \geq 8 \end{cases}$$

har invers $F^{-1}(u) = \min\{t : F(x) \geq u\}$

$$F^{-1}(u) = \begin{cases} 0 & \text{om } 0 < u \leq 0.5 \\ 16u - 8 & \text{om } 0.5 < u \leq 1 \end{cases}$$

Slumptalen $u_1 = 0.45$, $u_2 = 0.78$, $u_3 = 0.52$ och $u_4 = 0.34$ genererar utfallen $x_i = F_X^{-1}(u_i)$:

$$\begin{aligned} x_1 &= F_X^{-1}(0.45) = 0 \\ x_2 &= F_X^{-1}(0.78) = 16 \cdot 0.78 - 8 = 4.48 \\ x_3 &= F_X^{-1}(0.52) = 16 \cdot 0.52 - 8 = 0.32 \\ x_4 &= F_X^{-1}(0.27) = 0. \end{aligned}$$

10.1 Uppmätt kariesindex, x_1, \dots, x_n , för $n = 22$ personer ges nedan:

41 47 66 73 48 52 49 54 61 62 47
52 65 61 69 31 54 53 50 47 36 69

Det ordnade stickprovet $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ ges av

31 36 41 47 47 47 48 49 50 52 52
53 54 54 61 61 62 65 66 69 69 73

Datamängden kan sammanfattas av storheter som

Aritmetiskt medelvärde	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	=	53.995
Stickprovsvarians	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	=	118.71
Stickprovsstandardavvikelse	s	=	10.896
Variationskoefficient	s/\bar{x}	=	0.2019
Minsta värde	$x_{(1)}$	=	31
Största värde	$x_{(22)}$	=	73
Variationsbredd	$x_{(22)} - x_{(1)}$	=	42
Median	$\tilde{x}_{0.50} = \frac{x_{(10)} + x_{(11)}}{2}$	=	52.5
Undre kvartil	$\tilde{x}_{0.75}$	=	47
Övre kvartil	$\tilde{x}_{0.25}$	=	62
Interkvartilavstånd	$\tilde{x}_{0.75} - \tilde{x}_{0.25}$	=	15

10.2 Dataserien x_1, \dots, x_n , $n = 30$:

3 3 2 4 2 5 2 7 2 1 2 0 3 3 4
1 3 9 3 2 3 4 1 2 1 4 7 6 3 1

har medelvärde

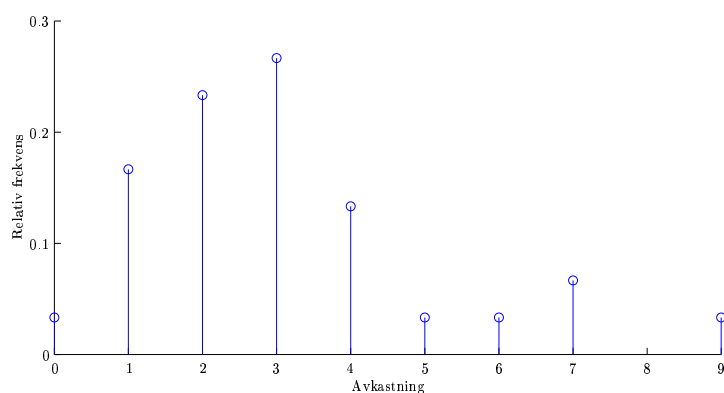
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 3.1$$

och stickprovsstandardavvikelse

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 2.0401.$$

Frekvenstabell:

Värde	Absolut frekvens	Relativ frekvens	Akkumulerad rel.
0	1	0.033333	0.033333
1	5	0.16667	0.2
2	7	0.23333	0.43333
3	8	0.26667	0.7
4	4	0.13333	0.83333
5	1	0.033333	0.86667
6	1	0.033333	0.9
7	2	0.066667	0.96667
8	0	0	0.96667
9	1	0.033333	1



10.3 Datamängden x_1, \dots, x_n , $n = 50$,

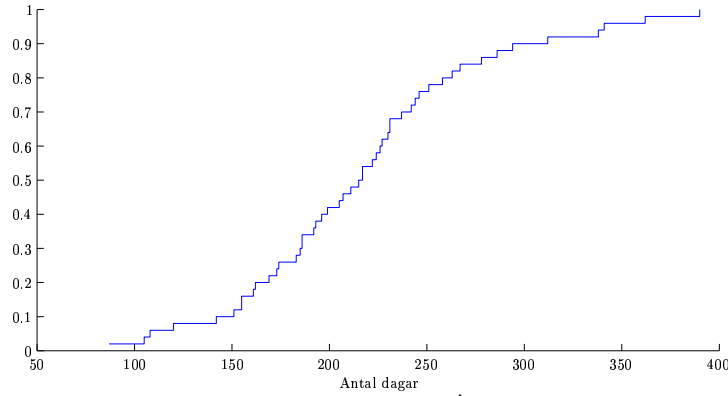
87 105 108 120 142 151 155 155 161 162
 169 173 174 183 185 186 186 192 193 196
 199 205 207 211 215 217 217 222 224 226
 227 230 231 231 237 242 244 246 251 258
 263 267 278 286 294 312 338 341 362 390

har medelvärde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 217.08$$

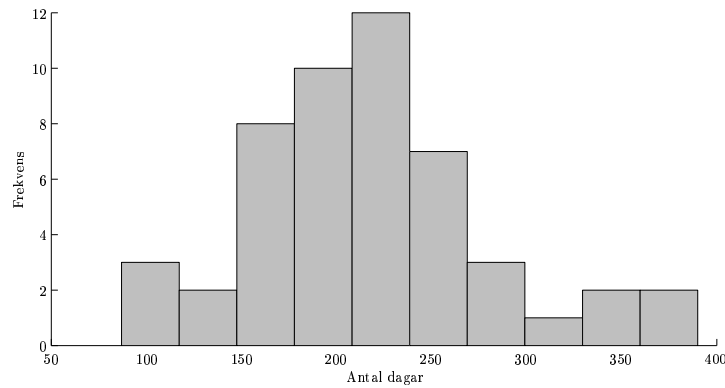
och stickprovsstandardavvikelse

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 64.25.$$



Den empiriska fördelningsfunktionen $\hat{F}(x)$:

$$\hat{F}(x) = \frac{1}{n} \cdot (\# \text{ observationer } \leq x)$$



Histogram över livlängden i antalet dagar för specialbatterier.

10.4 Låt x_1, \dots, x_n , $n = 10$, vara den första mätserien och y_1, \dots, y_m , $m = 5$, den andra mätserien. Sammanlagt till en mätserie om $m + n$ värden betecknar vi observationerna enligt

$$\begin{array}{c} x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m \\ \downarrow \\ w_1, w_2, \dots, w_n, w_{n+1}, w_{n+2}, \dots, w_{n+m} \end{array}$$

Då blir

$$\begin{aligned} \bar{w} &= \frac{1}{n+m} \sum_{i=1}^{n+m} w_i = \frac{1}{n+m} \left(\sum_{i=1}^n w_i + \sum_{i=n+1}^{n+m} w_i \right) = \frac{1}{n+m} \left(\sum_{i=1}^n x_i + \sum_{i=1}^m y_i \right) \\ &= \frac{1}{n+m} (n\bar{x} + m\bar{y}) = \frac{n\bar{x} + m\bar{y}}{n+m} = 5311.7. \end{aligned}$$

Stickprovsvariansen för första stickprovet är

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

så $\sum_{i=1}^n x_i^2 = (n-1)s_x^2 + n\bar{x}^2 = 2.8228 \cdot 10^8$. På samma sätt är $\sum_{i=1}^m y_i^2 = (m-1)s_y^2 + m\bar{y}^2 = 140927441$

$$\sum_{i=1}^{n+m} w_i^2 = \sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2 = 4.2321 \cdot 10^8.$$

Alltså,

$$s_w^2 = \frac{1}{n+m-1} \left(\sum_{i=1}^{n+m} w_i^2 - (n+m)\bar{w}^2 \right) = 23.764$$

$$\text{och } s_w = \sqrt{s_w^2} = 4.87.$$

10.5 Vi har $n = 800$ observationer med

$$\bar{x} = 9.496 \quad \text{och} \quad s_x = 0.345$$

Låt x_1 vara den felstämplade observationen, dvs $x_1 = 1.56$ istället för 9.56. Med y_1, \dots, y_n som korrekt observationsserie är

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \left(y_1 - x_1 + \sum_{i=1}^n x_i \right) = \frac{y_1 - x_1}{n} + \bar{x} = \frac{9.56 - 1.56}{800} + 9.496 = 9.506.$$

Nu är

$$s_x^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

så

$$\sum_{i=1}^n x_i^2 = (n-1)s_x^2 + n\bar{x}^2 = (800-1)(0.345)^2 + 800 \cdot (9.496)^2 = 72234$$

Alltså är

$$\sum_{i=1}^n y_i^2 = y_1^2 - x_1^2 + \sum_{i=1}^n x_i^2 = (9.56)^2 - (1.56)^2 + 72234 = 72323$$

och

$$s_y^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right) = \frac{1}{799} (72323 - 800 \cdot (9.506)^2) = 0.040106$$

och $s_y = 0.2003$.

10.6 De $n = 200$ observationerna x_1, \dots, x_n sammanfattas av klasstabellen:

Klassmitt (tkr)	Frekvens
17.5	5
22.5	19
27.5	79
32.5	33
37.5	18
42.5	15
47.5	10
52.5	5
57.5	4
62.5	5
67.5	3
72.5	4

Medelvärdet beräknas enligt:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \approx \frac{1}{n} (17.5 \cdot 5 + 22.5 \cdot 19 + \dots + 72.5 \cdot 4) = 34.225 \text{ tkr}$$

och stickprovsvariansen enligt:

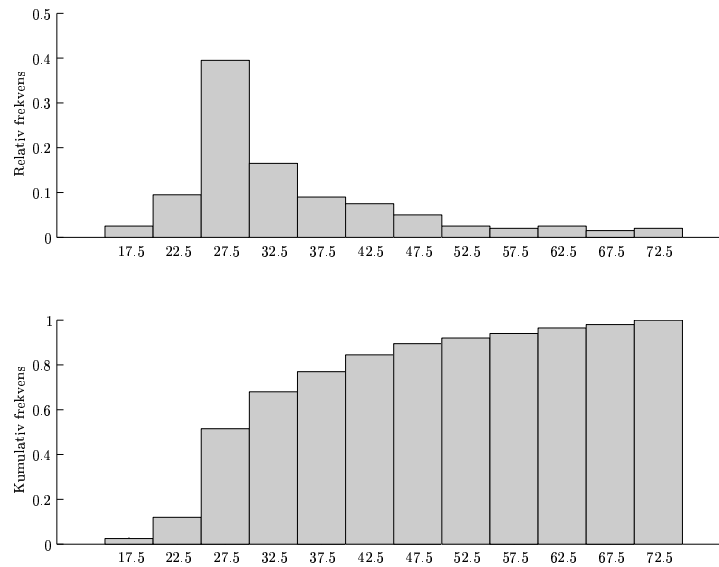
$$s_x^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \approx \frac{1}{n-1} ((17.5^2 \cdot 5 + 22.5^2 \cdot 19 + \dots + 72.5^2 \cdot 4) - n\bar{x}^2) = 139.09$$

så $s_x = 11.794$ (tkr).

Med den empiriska fördelningsfunktionen

$$\hat{F}(x) = \frac{1}{n} \cdot (\# \text{ observationer} \leq x)$$

fås att $\hat{F}(30.0) = 0.515$ och $\hat{F}(25.0) = 0.12$. Linjär interpolation ger $\hat{F}(29.81) = 0.50$ så medianen är approximativt 30000 kronor. Medelvärde ≈ 34000 och medianen ≈ 30000 skiljer sig åt. Fördelningen verkar inte vara symmetrisk. Kanske en lognormalfördelning beskriver data bättre.



Histogram över relativa frekvenser (överst) och kumulativa relativa frekvenser (nederst) för löneinkomsten per år i tusental kronor.

10.7 Låt x_1, \dots, x_n vara taxeringsvärdena och y_1, \dots, y_n motsvarande försäljningspris för $n = 10$ stycken villor.

Standardmässiga beräkningar ger

$$\bar{x} = 767, \bar{y} = 965 \quad s_x = 201.65, s_y = 231.67.$$

Kovariansens i datamaterialet är

$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) = \frac{1}{9} (7811000 - 10 \cdot 767 \cdot 965) = 45494$$

så korrelationen i datamaterialet är

$$r = \frac{c_{xy}}{s_x s_y} = 0.97384.$$

10.8 Observationer från poissonfördelningen kan skapas med funktionerna i avsnitt 8.6. MatLab-koden

```
>> n = 50;
>> my = 6;
>> x = poissrnd(my,1,n);
```

skapade $n = 50$ observationer x_1, \dots, x_n från Poissonfördelningen med väntevärde $\mu = 6$:

```
6 9 7 4 9 4 4 3 5 5
6 1 8 6 10 4 3 4 7 11
7 6 7 3 2 6 4 7 4 10
7 3 7 4 8 4 13 3 8 4
5 6 3 8 10 4 8 4 10 10
```

med medelvärde

```
>> mean(x)
```

```
ans =
```

```
6.0200
```

och stickprovsvarians

```
>> std(x)^2
```

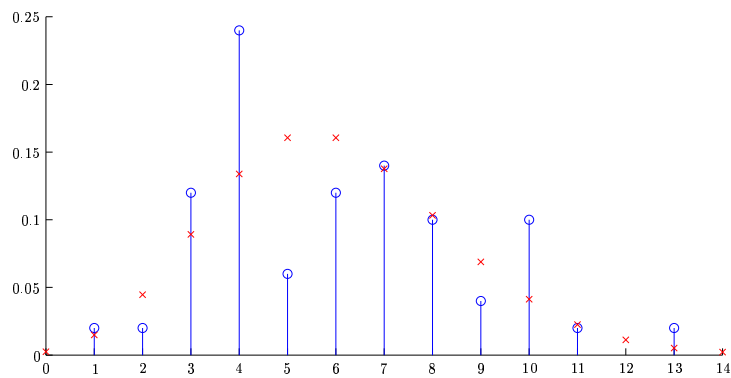
```
ans =
```

```
7.0404
```

som båda skall jämföras med Poissonfördelningens värde $\mu = 6$.

Ett stolpdigram för den relativa frekvensen plottas med

```
>> x0 = unique(x);
>> f = hist(x,x0);
>> stem(x0,f/n)
```



Ett stolpdigram med den empiriska fördelningen, markerad med cirklar, för $n = 50$ observationer från Po(6)-fördelningen där fördelningen är inritad med kryss.

10.9 Observationer från normalfördelningen kan skapas med funktionerna i avsnitt 8.6. MatLab-koden

```
>> n = 1000;
>> my = 3;
>> sigma = 1;
>> x = normrnd(my,sigma,1,n);
```

skapade $n = 1000$ observationer x_1, \dots, x_n från normalfördelningen (med parametrar $\mu = 3$ och $\sigma = 1$) med medelvärde

```
>> mean(x)
```

```
ans =
```

```
2.9569
```

och stickprovsstandardavvikelse

```
>> std(x)
```

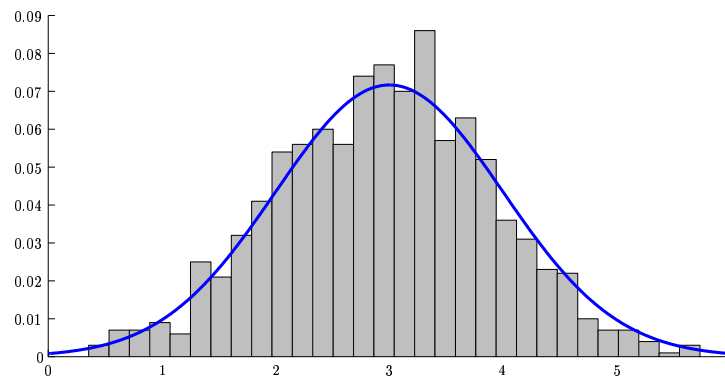
```
ans =
```

```
0.9435
```

Observationerna delas in i 30 klasser och ett histogram plottas med

```
>> [f,x0] = hist(x,30);
```

```
>> bar(x0,f,1)
```



Ett histogram med den empiriska fördelningen för $n = 1000$ observationer från $N(3, 1)$ -fördelningen där fördelningens täthet är inritad med heldragen linje.

11.1 Definitionen är att en skattning θ_{obs}^* är väntevärdesriktig skattning av en parameter θ om $E(\theta^*) = \theta$.

1. Nej, skattningen $\theta_{\text{obs}}^* = \theta^*(x_1, \dots, x_n)$ kan vara vilken funktion av x_1, \dots, x_n som helst, inte nödvändigtvis ett medelvärde.
2. Nej, inte heller ett viktat medelvärde. Tag som exempel skattningen s^2 av variansen för en stokastisk variabel.
3. Ja, en tolkning av väntvärdet $E(\theta^*)$ är att det är det genomsnittliga värdet av ett stort antal försök. (Stora talens lag säger att medelvärdet konvergerar mot väntevärdet.)
4. Nej, av flera skäl. Ett — det finns ingenting i väntevärdesriktigheten som säger att det är ett väntevärde som skattas. Två — skattningen θ_{obs}^* är ett utfall av en stokastisk variabel och ger olika skattningar i olika försök.
5. Nej, skattningen θ_{obs}^* är ett utfall av en stokastisk variabel och ger olika skattningar i olika försök.
6. Ja, detta är definitionen av väntevärdesriktighet.

11.2 Låt x_1, \dots, x_n vara batteriernas uppmätta livslängder. Vi modellerar dessa som utfall av oberoende och likafördelade stokastiska variabler X_1, \dots, X_n , med $\mu = E(X_i)$ och $\sigma^2 = V(X_i)$.

Väntevärdet μ skattas med

$$\mu_{\text{obs}}^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 5.2$$

och variansen σ^2 med

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i^2 \right) - n\bar{x}^2 \right) = 1.7,$$

det vill säga, standardavvikelsen σ skattas med $s = \sqrt{1.7} = 1.3038$.

Skattningen $\mu_{\text{obs}}^* = \bar{x}$ är ett utfall av stickprovsvariabeln

$$\mu^* = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

som har väntevärde

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu,$$

vilket visar att skattningen μ_{obs}^* är väntevärdesriktig, och varians

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n},$$

det vill säga $D(\bar{X}) = \sigma/\sqrt{n}$. Standardavvikelsen $D(\bar{X}) = \sigma/\sqrt{n}$ skattas med $s/\sqrt{n} = 0.5831$ och kallas för medelfelet för \bar{x} .

11.3 Låt X beskriva uppmätt halt av ämnet. Modell:

$$X = \text{halt} + \text{mätfel} = \mu + \epsilon$$

där ϵ är en stokastisk variabel med $E(\epsilon) = 0$ och $D(\epsilon) = \sigma = 0.5$. Då är

$$E(X) = E(\mu + \epsilon) = \mu + E(\epsilon) = \mu$$

och $V(X) = V(\mu + \epsilon) = V(\epsilon) = \sigma^2$. Skattningen \bar{x} av μ beskrivs av stickprovsvariabeln

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

som har

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu,$$

det vill säga \bar{x} är en väntevärdesriktig skattning av μ , och

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n \underbrace{V(X_i)}_{=\sigma^2} = \frac{\sigma^2}{n}.$$

Om n är stor så är $\sum_{i=1}^n X_i$ approximativt normalfördelad, det vill säga,

$$\bar{X} \text{ är approximativt } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\begin{aligned} P(|\bar{X} - \mu| \leq 0.25) &= P(-0.25 \leq \bar{X} - \mu \leq 0.25) = P\left(\frac{-0.25}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.25}{\sigma/\sqrt{n}}\right) \\ &\approx \Phi(0.5\sqrt{n}) - \Phi(-0.5\sqrt{n}) = 2\Phi(0.5\sqrt{n}) - 1 = 0.99. \end{aligned}$$

Alltså är $1 - \Phi(0.5\sqrt{n}) = 0.005$ men $1 - \Phi(\lambda_{0.005}) = 0.005$ där $\lambda_{0.005} = 2.5758$, vilket ger

$$0.50\sqrt{n} = \lambda_{0.005} \quad \text{eller} \quad n = \left(\frac{\lambda_{0.005}}{0.50}\right)^2 = 26.54,$$

alltså välj $n \geq 27$.

Om man inte kan förutsätta normalfördelning ger Tjebychovs olikhet en övre gräns för vad n behöver vara.

$$0.01 = P(|\bar{X} - \mu| > 0.25) \leq \frac{V(\bar{X})}{0.25^2} = \frac{\sigma^2/n}{0.25^2}$$

Vilket ger

$$n \leq \frac{\sigma^2}{0.25^2 \cdot 0.01} = 400,$$

det vill säga oavsett fördelning behöver inte mer än 400 mätningar göras.

Det är viktigt att se skillnaden på μ , parametern i fördelningen, \bar{x} , skattningen av parameterens värde, och \bar{X} , den stokastiska variabel som beskriver skattningen.

- 11.4 Låt X_1 och X_2 vara oberoende och $N(\mu, \sigma)$ -fördelade stokastiska variabler där $\sigma = 0.15$ och $\mu = g$. Väntevärdet skattas med $\bar{x} = \frac{1}{2}(x_1 + x_2)$ som beskrivs av $\bar{X} = \frac{1}{2}(X_1 + X_2)$ som är en normalfördelad stokastisk variabel med väntevärde

$$E(\bar{X}) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

och varians

$$V(\bar{X}) = V\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \{\text{oberoende}\} = \left(\frac{1}{2}\right)^2 V(X_1) + \left(\frac{1}{2}\right)^2 V(X_2) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{\sigma^2}{2}$$

det vill säga med standardavvikelse $D(\bar{X}) = \sigma/\sqrt{2}$. Alltså \bar{X} är $N(\mu, \sigma/\sqrt{2}) = N(g, 0.1061)$.

- 11.5 Låt X_1 och X_2 vara oberoende och binomialfördelade, X_i är $\text{Bin}(1, p)$, $i = 1, 2$, med observerade värden x_1 respektive x_2 . Då är $E(X_i) = p$ och $V(X_i) = p(1-p)$.

Låt

$$p_{\text{obs}}^* = x_1 \quad \text{och} \quad \hat{p}_{\text{obs}} = \frac{x_1 + x_2}{2}.$$

Dessa skattningar modelleras av

$$p^* = X_1 \quad \text{och} \quad \hat{p} = \frac{X_1 + X_2}{2}.$$

Möjliga värden på X_i är $\{0, 1\}$ så möjliga värden på p^* ges av $\{0, 1\}$ medan för \hat{p} är det $\{0, 1/2, 1\}$.

b) Vi har att

$$E(p^*) = E(X_1) = p \quad E(\hat{p}) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{p}{2} + \frac{p}{2} = p$$

så båda skattningarna p_{obs}^* och \hat{p}_{obs} är väntevärdesriktiga.

c) Vidare så är, eftersom X_1 och X_2 är oberoende,

$$V(p^*) = V(X_1) = p(1-p) \quad V(\hat{p}) = V\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}V(X_1) + \frac{1}{4}V(X_2) = \frac{p(1-p)}{2}.$$

Eftersom $V(\hat{p}) < V(p^*)$ för alla $p \in (0, 1)$ är skattningen \hat{p}_{obs} effektivare än skattningen p_{obs}^* .

11.6 Låt x_1, \dots, x_n vara utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n . Väntevärdet μ skattas dels med

$$\mu_{\text{obs}}^* = \bar{x} \quad \text{och dels med} \quad \hat{\mu}_{\text{obs}} = \frac{x_1 + x_2}{2}.$$

a) Båda dessa skattningar är väntevärdesriktiga eftersom

$$E(\mu^*) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$

och

$$E(\hat{\mu}) = E\left(\frac{1}{2}(X_1 + X_n)\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_n) = \frac{\mu}{2} + \frac{\mu}{2} = \mu.$$

b) Eftersom X_1, \dots, X_n är oberoende så är

$$V(\mu^*) = V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

och

$$V(\hat{\mu}) = V\left(\frac{1}{2}(X_1 + X_n)\right) = \frac{1}{4}V(X_1) + \frac{1}{4}V(X_n) = \frac{\sigma^2}{2}.$$

Om $n > 2$ är $V(\mu^*) < V(\hat{\mu})$ och skattningen μ_{obs}^* är effektivare än skattningen $\hat{\mu}_{\text{obs}}$. Notera även att $V(\mu^*) \rightarrow 0$ då $n \rightarrow \infty$ medan $V(\hat{\mu})$ är konstant, så skattningen μ_{obs}^* är en konsistent skattning av μ men $\hat{\mu}$ är det inte.

11.7 Den stokastiska variabeln X är binomialfördelad, X är $\text{Bin}(n, p)$ och Y är hypergeometriskt fördelad, Y är $\text{Hyp}(N, n, p)$. Låt $p^* = X/n$ och $\hat{p} = Y/n$.

a) Då är

$$E(p^*) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$$

och

$$V(p^*) = V\left(\frac{1}{n}X\right) = \frac{1}{n^2}V(X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

Vidare så är

$$E(\hat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{1}{n}np = p$$

och

$$\begin{aligned} V(\hat{p}) &= V\left(\frac{1}{n}Y\right) = \frac{1}{n^2}V(Y) = \frac{1}{n^2}np(1-p)\frac{N-n}{N-1} = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1} = V(p^*) \underbrace{\frac{N-n}{N-1}}_{\leq 1} \\ &\leq V(p^*), \end{aligned}$$

med likhet endast om $n = 1$.

b) För $n > 1$ är $V(\hat{p}) < V(p^*)$ och skattningen \hat{p}_{obs} är effektivare än skattningen p_{obs}^* .

c) Med observationer erhålls skattningarna $p_{\text{obs}}^* = \hat{p}_{\text{obs}} = 23/100 = 0.23$ av andelen.

Stickprovsvariablernas varianser skattas med hjälp av skattningarna av p . Det vill säga,

$$V(p^*) = \frac{p(1-p)}{n} \quad \text{skattas med} \quad \frac{0.23 \cdot (1 - 0.23)}{100} = 0.001771$$

och

$$V(\hat{p}) = \frac{p(1-p)}{n} \frac{N-n}{N-1} \quad \text{skattas med} \quad \frac{0.23 \cdot (1-0.23)}{100} \frac{1000-100}{1000-1} = 0.00160.$$

Detta medför att standardavvikelserna $D(p^*)$ och $D(\hat{p})$ skattas med 0.0421 respektive 0.0399. Dessa skattningar kallas för medelfelen för p_{obs}^* respektive \hat{p}_{obs} .

11.8 Att θ_{obs}^* och $\hat{\theta}_{\text{obs}}$ är väntevärdesriktiga skattningar av en parameter θ betyder att motsvarande stickprovsvariabler uppfyller $E(\theta^*) = \theta$ och $E(\hat{\theta}) = \theta$. Med skattningen

$$\tilde{\theta}_{\text{obs}} = a\theta_{\text{obs}}^* + (1-a)\hat{\theta}_{\text{obs}}$$

så är

$$E(\tilde{\theta}) = E(a\theta^* + (1-a)\hat{\theta}) = a \underbrace{E(\theta^*)}_{=\theta} + (1-a) \underbrace{E(\hat{\theta})}_{=\theta} = \theta$$

för alla värden a , så $\tilde{\theta}_{\text{obs}}$ är en väntevärdesriktig skattning av θ oavsett värdet på a . Vidare så är

$$\begin{aligned} V(\tilde{\theta}) &= V(a\theta^* + (1-a)\hat{\theta}) = \{\text{oberoende}\} = a^2 \underbrace{V(\theta^*)}_{=\sigma_1^2} + (1-a)^2 \underbrace{V(\hat{\theta})}_{=\sigma_2^2} \\ &= a^2\sigma_1^2 + (1-a)^2\sigma_2^2 = g(a), \end{aligned}$$

en funktion av a . Den effektivaste skattningen erhålls om a väljes så att variansen $g(a)$ minimeras. Detta a bestäms som lösningen till

$$0 = \frac{d}{da}g(a) = \sigma_1^2 \cdot 2a + \sigma_2^2 \cdot 2(1-a)(-1) = 2[a(\sigma_1^2 + \sigma_2^2) - \sigma_2^2],$$

där lösningen är

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Teckenstudium av andraderivatans visar att det är ett minimum som vi fått fram. Alltså, den effektivaste skattningen fås om

$$\tilde{\theta}_{\text{obs}} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_{\text{obs}}^* + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \hat{\theta}_{\text{obs}}.$$

11.9 Låt x_1, \dots, x_n vara utfallen av de stokastiska variablerna X_1, \dots, X_n som antas vara oberoende och $N(\mu, \sigma)$ -fördelade.

a) Om μ är känd så kan variansen σ^2 skattas med $(\hat{\sigma}^2)_{\text{obs}}$ där

$$(\hat{\sigma}^2)_{\text{obs}} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

Denna skattning är väntevärdesriktig eftersom

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right) = \frac{1}{n} \sum_{i=1}^n \underbrace{E((X_i - \mu)^2)}_{=\sigma^2} = \frac{1}{n} n\sigma^2 = \sigma^2.$$

Med de $n = 4$ observationerna x_1, \dots, x_4 och $\mu = 1457.0$ så är

$$(\hat{\sigma}^2)_{\text{obs}} = \frac{1}{n} \sum_{i=1}^n (x_i - 1457.0)^2 = 0.8175.$$

b) Om μ är okänd måste väntevärdet skattas. En väntevärdesriktig skattning av σ^2 är s^2 där

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

vilket, då μ skattas med $\bar{x} = 1457.4$, ger skattningen $s^2 = 0.84917$.

Att s^2 är väntevärdesriktig ges av följande uträkning.

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = E\left(\frac{1}{n-1} \sum_{i=1}^n X_i^2 - 2\bar{X}X_i + \bar{X}^2\right) \\ &= E\left(\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2\right) - 2\bar{X} \left(\sum_{i=1}^n X_i\right) + n\bar{X}^2\right) \\ &= E\left(\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2\right) - 2\bar{X}n\bar{X} + n\bar{X}^2\right) = E\left(\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2\right) - n\bar{X}^2\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right) = \frac{1}{n-1} \left(\sum_{i=1}^n (\mu^2 + \sigma^2) - n(\mu^2 + \frac{\sigma^2}{n})\right) \\ &= \frac{1}{n-1} (n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) = \frac{1}{n-1} (n-1)\sigma^2 = \sigma^2. \end{aligned}$$

c) Nej, av uträkningarna ovan framgår att endast $E(X_i)$ och $V(X_i)$ har utnyttjats.

11.10 Den stokastiska variabeln X har sannolikhetsfunktion $p_X(k) = (1-\theta)^{k-1}\theta$ för $k = 1, 2, 3, \dots$ där $0 < \theta < 1$. Låt x_1, \dots, x_n vara ett stickprov av X .

a) Då observationerna är utfall av oberoende stokastiska variabler blir Likelihoodfunktionen

$$\begin{aligned} L(\theta) &= p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n) = (1-\theta)^{x_1-1}\theta \cdots (1-\theta)^{x_n-1}\theta \\ &= \theta^n (1-\theta)^{\sum_{i=1}^n x_i - n}. \end{aligned}$$

b) Vi söker det värde på θ som maximerar $L(\theta)$. Det är samma värde som maximerar den logaritmerade likelihoodfunktionen

$$\ln(L(\theta)) = n \ln(\theta) + \ln(1-\theta) \left(\sum_{i=1}^n x_i - n\right).$$

Detta maximum bestäms som lösningen till

$$\begin{aligned} 0 &= \frac{d}{d\theta} \ln(L(\theta)) = \frac{d}{d\theta} \left[n \ln(\theta) + \ln(1-\theta) \left(\sum_{i=1}^n x_i - n\right) \right] = \frac{n}{\theta} - \frac{1}{1-\theta} \underbrace{\left(\sum_{i=1}^n x_i - n\right)}_{=n\bar{x}} \\ &= \frac{n}{\theta(1-\theta)} (1-\theta\bar{x}). \end{aligned}$$

Alltså är $\theta = 1/\bar{x}$ det värde som ger maximum. (Kontroll av andraderivatans ger att det verkligen är ett maximum som erhållits.) Alltså, ML-skattningen av θ är

$$\theta_{\text{obs}}^* = \frac{1}{\bar{x}}.$$

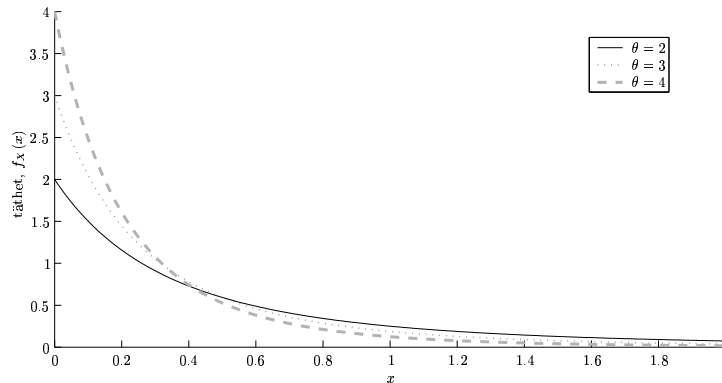
Med $(x_1, \dots, x_n) = (4, 5, 4, 6, 4, 1)$ är $\bar{x} = 4$ och $\theta_{\text{obs}}^* = 1/\bar{x} = 1/4$.

Notera att X är ffg(θ)-fördelad och har väntevärde $E(X) = 1/\theta$. Om detta väntevärde skattas med \bar{x} fås $\theta_{\text{obs}}^* = 1/\bar{x}$ som skattning av θ .

11.11 Den stokastiska variabeln X har täthetsfunktion

$$f_X(x) = \theta(1+x)^{-(\theta+1)}, \quad x \geq 0,$$

där möjliga värden på parametern θ är 2, 3 eller 4. De tre möjliga tätheterna för X visas i figuren nedan.



De tre olika tätheterna motsvarande parametervärdena $\theta = 2, 3, 4$.

Med stickprovet x_1, x_2 blir likelihoodfunktionen

$$L(\theta) = p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) = \theta(1+x_1)^{-(\theta+1)}\theta(1+x_2)^{-(\theta+1)} = \theta^2 [(1+x_1)(1+x_2)]^{-(\theta+1)}.$$

Med observationerna $(x_1, x_2) = (0.2, 0.8)$ blir $L(\theta) = \theta^2 \cdot (2.16)^{-(\theta+1)}$ och

Parameter	$L(\theta)$
$\theta = 2$	0.39692
$\theta = 3$	0.41345
$\theta = 4$	0.34029

så valet $\theta = 3$ ger maximum. Alltså, ML-skattningen av θ är $\theta_{\text{obs}}^* = 3$.

11.12 Låt x_1, \dots, x_n vara antalet telefonsamtal under olika dagar. Vi ansätter modellen att x_1, \dots, x_n är utfall av oberoende Poisson(μ)-fördelade stokastiska variabler X_1, \dots, X_n .

Maximum likelihoodskattningen av μ är det värde på μ som maximerar

$$L(\mu) = p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n) = \frac{\mu^{x_1}}{x_1!} e^{-\mu} \cdots \frac{\mu^{x_n}}{x_n!} e^{-\mu} = \frac{\mu^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!} e^{-n\mu}.$$

Det är samma μ som maximerar

$$\ln(L(\mu)) = \ln\left(\frac{\mu^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!} e^{-n\mu}\right) = \ln(\mu) \sum_{i=1}^n x_i - n\mu - \sum_{i=1}^n \ln(x_i!).$$

Detta maximum bestäms som det μ som löser

$$0 = \frac{d}{d\mu} \ln(L(\mu)) = \frac{d}{d\mu} \left[\ln(\mu) \sum_{i=1}^n x_i - n\mu - \sum_{i=1}^n \ln(x_i!) \right] = \frac{1}{\mu} \sum_{i=1}^n x_i - n$$

vilket ger $\mu = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$. Kontroll av andraderivatan ger att detta är ett maximum. Skattningen $\mu_{\text{obs}}^* = \bar{x}$ av μ beskrivs av stickprovsvariabeln $\mu^* = \bar{X}$ som har

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu,$$

vilket ger att \bar{x} är en väntevärdesriktig skattning av μ , och

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n \underbrace{V(X_i)}_{=\mu} = \frac{\mu}{n},$$

det vill säga $D(\bar{X}) = \sqrt{\mu/n}$.

Med observationer x_1, \dots, x_8 :

$$115 \quad 82 \quad 108 \quad 106 \quad 118 \quad 87 \quad 99 \quad 92$$

fås skattningen $\bar{x} = 100.88$. En skattning av $D(\bar{X}) = \sqrt{\mu/n}$ ges av

$$\sqrt{\frac{\mu_{\text{obs}}^*}{n}} = \sqrt{\frac{100.88}{8}} = 3.551$$

och kallas för skattningens medelfel.

11.13 Låt x_1, \dots, x_n vara de uppmätta tiderna mellan fel hos den komplicerade tekniska utrustningen. Med modellen att x_1, \dots, x_n är utfall av oberoende exponential(λ)-fördelade stokastiska variabler X_1, \dots, X_n är likelihoodfunktionen

$$L(\lambda) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \lambda e^{-\lambda x_1} \cdots \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

Det värde på λ som maximerar $L(\lambda)$ är samma som maximerar

$$\ln(L(\lambda)) = \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^n x_i}\right) = n \ln(\lambda) - \lambda \underbrace{\sum_{i=1}^n x_i}_{=n\bar{x}} = n \ln(\lambda) - \lambda n \bar{x}.$$

Maximum bestäms som det λ som löser

$$0 = \frac{d}{d\lambda} \ln(L(\lambda)) = \frac{d}{d\lambda} [n \ln(\lambda) - \lambda n \bar{x}] = \frac{n}{\lambda} - n \bar{x} = n \left(\frac{1}{\lambda} - \bar{x}\right),$$

det vill säga $\lambda = 1/\bar{x}$. Kontroll av andraderivatans ger att detta är ett maximum. Alltså, ML-skattningen av λ är $\lambda_{\text{obs}}^* = 1/\bar{x}$.

Funktionen $g(x) = 1/x$ har derivata $g'(x) = -1/x^2$. Enligt felfortplantningsformeln för varianser har vi att

$$V(\lambda^*) = V(g(\bar{X})) \approx [g'(E(\bar{X}))]^2 V(\bar{X}) = [g'(1/\lambda)]^2 \frac{1/\lambda^2}{n} = \left[\frac{-1}{(1/\lambda)^2}\right]^2 \frac{1/\lambda^2}{n} = \frac{\lambda^2}{n}.$$

Medelfelet för skattningen är således

$$d(\lambda^*) = \sqrt{\frac{(\lambda_{\text{obs}}^*)^2}{n}} = \frac{1}{\bar{x}\sqrt{n}}.$$

Anmärkning. En exakt, och ganska komplicerad, uträkning ger att standardavvikelsen för λ^* är $D(\lambda^*) = n\lambda/((n-1)\sqrt{n-2})$. Utnyttjandes detta fås medelfelet

$$d_2(\lambda^*) = \frac{n}{\bar{x}(n-1)\sqrt{n-2}}.$$

Kvoten mellan dessa medelfel beror inte av \bar{x} utan bara av antalet observationer n . Kvoten är här illustrerad för några stickprovsstorlekar

n	10	20	30	40	50	60	70	80	90	100
$d(\lambda^*)/d_2(\lambda^*)$	0.805	0.901	0.934	0.950	0.960	0.967	0.972	0.975	0.978	0.980

11.14 Låt x_1, \dots, x_n vara utfall av X_1, \dots, X_n där de stokastiska variablerna är oberoende och har täthetsfunktion

$$f_X(x) = \theta x^{\theta-1}, \quad 0 < x < 1.$$

Likelihoodfunktionen blir

$$L(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \{\text{ober.}\} = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \theta x_1^{\theta-1} \cdots \theta x_n^{\theta-1} = \theta^n (x_1 \cdots x_n)^{\theta-1}.$$

Det är samma θ som maximerar

$$\ln(L(\theta)) = \ln\left(\theta^n (x_1 \cdots x_n)^{\theta-1}\right) = n \ln(\theta) + (\theta - 1) \sum_{i=1}^n \ln(x_i).$$

Maximum bestäms som det θ som löser

$$0 = \frac{d}{d\theta} \ln(L(\theta)) = \frac{d}{d\theta} \left[n \ln(\theta) + (\theta - 1) \sum_{i=1}^n \ln(x_i) \right] = \frac{n}{\theta} + \sum_{i=1}^n \ln(x_i)$$

det vill säga $\theta = -n / \sum_{i=1}^n \ln(x_i)$. Kontroll av andraderivatatan ger att detta är ett maximum. (Notera att $\ln(x_i) < 0$ för $0 < x_i < 1$.) Alltså, maximum-likelihoodmetodens skattning av θ är $\theta_{\text{obs}}^* = -n / \sum_{i=1}^n \ln(x_i)$.

11.15 Låt x_1, \dots, x_n vara utfall av X_1, \dots, X_n där de stokastiska variablerna är oberoende och Rayleigh(a)-fördelade, det vill säga har täthetsfunktion

$$f_X(x) = (x/a)e^{-x^2/2a} \quad x \geq 0.$$

Likelihoodfunktionen blir

$$\begin{aligned} L(a) &= f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \{\text{ober.}\} = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \frac{x_1}{a} e^{-x_1^2/2a} \cdots \frac{x_n}{a} e^{-x_n^2/2a} \\ &= \frac{x_1 \cdots x_n}{a^n} e^{-\frac{1}{2a} \sum_{i=1}^n x_i^2}. \end{aligned}$$

Vi söker det värde på a som maximerar $L(a)$, och det är samma a som maximerar

$$\ln(L(a)) = \ln\left(\frac{x_1 \cdots x_n}{a^n} e^{-\frac{1}{2a} \sum_{i=1}^n x_i^2}\right) = -n \ln(a) + \sum_{i=1}^n \ln(x_i) - \frac{1}{2a} \sum_{i=1}^n x_i^2.$$

Detta maximum bestäms som det a som löser

$$0 = \frac{d}{da} \ln(L(a)) = \frac{d}{da} \left[-n \ln(a) + \sum_{i=1}^n \ln(x_i) - \frac{1}{2a} \sum_{i=1}^n x_i^2 \right] = -\frac{n}{a} + \frac{1}{2a^2} \sum_{i=1}^n x_i^2 = \frac{-1}{a} \left[n - \frac{1}{2a} \sum_{i=1}^n x_i^2 \right]$$

vilket medför att $a = \frac{1}{2n} \sum_{i=1}^n x_i^2$. Kontroll av andraderivatatan ger att detta är ett maximum. Således, maximum-likelihoodskattningen av a är $a_{\text{obs}}^* = \frac{1}{2n} \sum_{i=1}^n x_i^2$.

11.16 Låt X_1, \dots, X_n vara oberoende och likformigt fördelade på intervallet $[-\theta, \theta]$, $\theta > 0$, med observerade värden x_1, \dots, x_n , det vill säga de stokastiska variablerna har täthetsfunktion

$$f_X(x) = \frac{1}{2\theta} \quad \text{för } -\theta \leq x \leq \theta.$$

Maximum likelihoodskattningen är det värde på θ som maximerar

$$L(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \{\text{oberoende}\} = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \frac{1}{2\theta} = (2\theta)^{-n}$$

för $-\theta \leq x_1, \dots, x_n \leq \theta$. Eftersom $L(\theta)$ är avtagande i θ skall man göra θ så liten som möjligt. Kravet $-\theta \leq x_1, \dots, x_n \leq \theta$, det vill säga $|x_1|, \dots, |x_n| \leq \theta$, gör att det minsta möjliga värdet på θ ges av $\theta = \max(|x_1|, \dots, |x_n|)$. Alltså, maximum-likelihoodskattningen av θ är $\theta_{\text{obs}}^* = \max(|x_1|, \dots, |x_n|)$.

Eftersom denna skattning omöjlig kan överskatta θ kommer den ha ett systematiskt fel, $E(\theta^*) < \theta$.

För att bestämma det systematiska felet beräknar vi först fördelningsfunktionen till θ^* . För $t \in [0, \theta]$ är

$$\begin{aligned} F_{\theta^*}(t) &= P(\theta^* \leq t) = P(\max(|X_1|, \dots, |X_n|) \leq t) = P(|X_1| \leq t, \dots, |X_n| \leq t) = \{\text{oberoende}\} \\ &= P(|X_1| \leq t) \cdots P(|X_n| \leq t) = P(|X| \leq t)^n = P(-t \leq X \leq t)^n = (F_X(t) - F_X(-t))^n. \end{aligned}$$

Alltså är

$$f_{\theta^*}(t) = \frac{d}{dt} F_{\theta^*}(t) = n(F_X(t) - F_X(-t))^{n-1} (f_X(t) + f_X(-t)).$$

I vårt fall är $f_X(-t) = f_X(t) = 1/(2\theta)$ och

$$F_X(t) = \int_{-\theta}^t f_X(x) dx = \frac{t - (-\theta)}{2\theta} = \frac{t + \theta}{2\theta}, \quad F_X(-t) = \frac{-t + \theta}{2\theta}$$

så

$$f_{\theta^*}(t) = n \left(\frac{\theta + t}{2\theta} - \frac{\theta - t}{2\theta} \right)^{n-1} \left(\frac{1}{2\theta} + \frac{1}{2\theta} \right) = n \left(\frac{t}{\theta} \right)^{n-1} \frac{1}{\theta}.$$

Slutligen får vi att

$$E(\theta^*) = \int_{-\infty}^{\infty} t f_{\theta^*}(t) dt = \int_0^{\theta} t \cdot n \left(\frac{t}{\theta} \right)^{n-1} \frac{1}{\theta} dt = \int_0^1 \theta u \cdot n u^{n-1} du = \theta n \left[\frac{u^{n+1}}{n+1} \right]_0^1 = \frac{n}{n+1} \theta.$$

Alltså är

$$E(\theta^*) = \frac{n}{n+1} \theta < \theta$$

och skattningen θ_{obs}^* är inte väntevärdesriktig. Dock, skattningen

$$\hat{\theta}_{\text{obs}} = \frac{n+1}{n} \theta_{\text{obs}}^* = \frac{n+1}{n} \max(|x_1|, \dots, |x_n|)$$

är sådan att

$$E(\hat{\theta}) = E\left(\frac{n+1}{n} \theta^*\right) = \frac{n+1}{n} E(\theta^*) = \frac{n+1}{n} \cdot \frac{n}{n+1} \theta = \theta,$$

dvs är väntevärdesriktig.

11.17 Den stokastiska variabeln X har täthetsfunktion

$$f_X(x) = \theta(1+x)^{-(\theta+1)}, \quad x \geq 0,$$

där möjliga värden på parametern θ är 2, 3 eller 4. Nu är

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \cdot \frac{\theta}{(1+x)^{\theta+1}} dx = \left[\frac{-x}{(1+x)^\theta} \right]_0^{\infty} + \int_0^{\infty} \frac{1}{(1+x)^\theta} dx \\ &= \left[\frac{-1/(\theta-1)}{(1+x)^{\theta-1}} \right]_0^{\infty} = \frac{1}{\theta-1}. \end{aligned}$$

Med $n = 2$ observationer $(x_1, x_2) = (0.2, 0.8)$ är minsta-kvadratskattningen av θ är det värde som minimerar

$$Q(\theta) = \sum_{i=1}^n (x_i - E(X_i))^2 = \left(0.2 - \frac{1}{\theta-1}\right)^2 + \left(0.8 - \frac{1}{\theta-1}\right)^2$$

För de olika värdena på θ erhålles

Parameter	$Q(\theta)$
$\theta = 2$	0.68
$\theta = 3$	0.18
$\theta = 4$	0.24

så valet $\theta = 3$ ger minimum. Alltså, MK-skattningen av θ är $\theta_{\text{obs}}^* = 3$.

11.18 Låt X_1, X_2, X_3 beskriva mätningarna på vinkeln AOC och X_4, X_5 på vinkeln AOB. Modell: X_1, \dots, X_n är oberoende och

$$E(X_1) = E(X_2) = E(X_3) = \theta_1 + \theta_2 \quad E(X_4) = E(X_5) = \theta_1$$

och $V(X_i) = \sigma^2$. Minsta kvadratmetodens skattning av (θ_1, θ_2) är det värde på (θ_1, θ_2) som minimerar

$$Q(\theta_1, \theta_2) = \sum_{i=1}^n (x_i - E(X_i))^2 = (x_1 - (\theta_1 + \theta_2))^2 + (x_2 - (\theta_1 + \theta_2))^2 + (x_3 - (\theta_1 + \theta_2))^2 \\ + (x_4 - \theta_1)^2 + (x_5 - \theta_1)^2.$$

Derivering med avseende på θ_1 och θ_2 ger

$$\begin{aligned} \frac{\partial}{\partial \theta_1} Q(\theta_1, \theta_2) &= -2(x_1 - (\theta_1 + \theta_2)) - 2(x_2 - (\theta_1 + \theta_2)) - 2(x_3 - (\theta_1 + \theta_2)) \\ &\quad - 2(x_4 - \theta_1) - 2(x_5 - \theta_1) \\ &= -2 \left[\sum_{i=1}^5 x_i - 5\theta_1 - 3\theta_2 \right] \\ \frac{\partial}{\partial \theta_2} Q(\theta_1, \theta_2) &= -2(x_1 - (\theta_1 + \theta_2)) - 2(x_2 - (\theta_1 + \theta_2)) - 2(x_3 - (\theta_1 + \theta_2)) \\ &= -2[x_1 + x_2 + x_3 - 3\theta_1 - 3\theta_2] \end{aligned}$$

Sätts derivatorna till 0 fås ekvationssystemet

$$\begin{cases} 5\theta_1 + 3\theta_2 = x_1 + x_2 + x_3 + x_4 + x_5 \\ 3\theta_1 + 3\theta_2 = x_1 + x_2 + x_3 \end{cases}$$

med lösningen

$$(\theta_1^*)_{\text{obs}} = \frac{x_4 + x_5}{2} \quad (\theta_2^*)_{\text{obs}} = \frac{x_1 + x_2 + x_3}{3} - (\theta_1^*)_{\text{obs}}.$$

b) Skattningen $(\theta_1^*)_{\text{obs}}$ är väntevärdesriktig ty

$$E(\theta_1^*) = E\left(\frac{1}{2}X_4 + \frac{1}{2}X_5\right) = \frac{1}{2}E(X_4) + \frac{1}{2}E(X_5) = \frac{1}{2}\theta_1 + \frac{1}{2}\theta_1 = \theta_1.$$

Vidare så är

$$\begin{aligned} E(\theta_2^*) &= E\left(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 - \theta_1^*\right) = \frac{1}{3}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{3}E(X_3) - E(\theta_1^*) \\ &= \frac{1}{3}(\theta_1 + \theta_2) + \frac{1}{3}(\theta_1 + \theta_2) + \frac{1}{3}(\theta_1 + \theta_2) - \theta_1 = \theta_2 \end{aligned}$$

så även $(\theta_2^*)_{\text{obs}}$ är väntevärdesriktig.

c) Skattningarna $(\theta_1^*)_{\text{obs}}$ och $(\theta_2^*)_{\text{obs}}$ är utfall av stokastiska variabler med varians

$$V(\theta_1^*) = V\left(\frac{1}{2}X_4 + \frac{1}{2}X_5\right) = \{\text{oberoende}\} = \frac{1}{2^2}\sigma^2 + \frac{1}{2^2}\sigma^2 = \frac{\sigma^2}{2}.$$

respektive

$$\begin{aligned} V(\theta_2^*) &= V\left(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 - \theta_1^*\right) = \{\text{ober.}\} = \frac{1}{3^2}\sigma^2 + \frac{1}{3^2}\sigma^2 + \frac{1}{3^2}\sigma^2 + (-1)^2V(\theta_1^*) = \frac{\sigma^2}{3} + \frac{\sigma^2}{2} \\ &= \frac{5}{6}\sigma^2. \end{aligned}$$

11.19 Låt x_1, x_2, x_3 vara mätningar på vinkeln vid B och x_4, \dots, x_7 de vid C. Minsta-kvadratmetodens skattning av θ är det värde på θ som minimerar

$$\begin{aligned} Q(\theta) &= \sum_{i=1}^7 (x_i - E(X_i))^2 = \sum_{i=1}^3 (x_i - \theta)^2 + \sum_{i=4}^7 (x_i - (90 - \theta))^2 \\ &= (x_1 - \theta)^2 + (x_2 - \theta)^2 + (x_3 - \theta)^2 \\ &\quad + (x_4 - (90 - \theta))^2 + (x_5 - (90 - \theta))^2 + (x_6 - (90 - \theta))^2 + (x_7 - (90 - \theta))^2 \end{aligned}$$

Funktionens maximipunkt hittas genom att derivatan sätts till noll.

$$\begin{aligned} \frac{d}{d\theta} Q(\theta) &= -2 \sum_{i=1}^3 (x_i - \theta) - 2 \sum_{i=4}^7 (90 - x_i - \theta) \\ &= -2(x_1 + x_2 + x_3 + (90 - x_4) + \dots + (90 - x_7) - 7\theta) = 0. \end{aligned}$$

Detta ger

$$\theta_{\text{obs}}^* = \frac{x_1 + x_2 + x_3 + (90 - x_4) + \dots + (90 - x_7)}{7},$$

eller, med siffror, $\theta_{\text{obs}}^* = 61.17$. Minsta-kvadratskattningen av vinkeln vid C blir då $90 - 61.17 = 28.83$.

Skattningen beskrivs av stickprovsvariabeln

$$\theta^* = \frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}.$$

Den har väntevärde

$$\begin{aligned} E(\theta^*) &= E\left(\frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}\right) \\ &= \frac{1}{7}(\underbrace{E(X_1)}_{=\theta} + \dots + \underbrace{E(X_3)}_{=\theta} + \underbrace{(90 - E(X_4))}_{=90-\theta} + \dots + \underbrace{(90 - E(X_7))}_{90-\theta}) = \frac{1}{7} \cdot 7\theta = \theta, \end{aligned}$$

vilket innebär att skattningen θ_{obs}^* är väntevärdesriktig. Vidare,

$$\begin{aligned} V(\theta^*) &= V\left(\frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}\right) \\ &= \frac{1}{7^2}(\underbrace{V(X_1)}_{=\sigma^2} + \dots + \underbrace{V(X_3)}_{=\sigma^2} + (-1)^2 \underbrace{V(X_4)}_{=\sigma^2} + \dots + (-1)^2 \underbrace{V(X_7)}_{=\sigma^2}) = \frac{\sigma^2}{7} \end{aligned}$$

så $D(\theta^*) = \sigma/\sqrt{7} = 0.1/\sqrt{7} = 0.0378$. Skattningen $90 - \theta_{\text{obs}}^*$ är ett utfall från en stokastisk variabel med samma standardavvikelse.

11.20 Låt x_1, \dots, x_n vara mätningarna av våglängden θ . Dessa mätvärden är utfall av de oberoende stokastiska variablerna X_1, \dots, X_n där $E(X_i) = \theta$. Minsta-kvadratskattningen av θ är det värde på θ som minimerar

$$Q(\theta) = \sum_{i=1}^n \frac{1}{V(X_i)} (x_i - E(X_i))^2.$$

Med mätningarna

x_i	79.1	80.0	81.3	81.9	81.7
$100D(X_i)$	2	1	2	3	1

fås

$$\frac{1}{10000} Q(\theta) = \frac{1}{2^2}(x_1 - \theta)^2 + \frac{1}{1^2}(x_2 - \theta)^2 + \frac{1}{2^2}(x_3 - \theta)^2 + \frac{1}{3^2}(x_4 - \theta)^2 + \frac{1}{1^2}(x_5 - \theta)^2$$

Lösning av

$$0 = \frac{d}{d\theta} \left[\frac{1}{10000} Q(\theta) \right] = 2 \left[\frac{1}{4}x_1 + x_2 + \frac{1}{4}x_3 + \frac{1}{9}x_4 + x_5 - \frac{47}{18}\theta \right] = 2 \left[210.9 - \frac{47}{18}\theta \right]$$

ger skattningen $\theta_{\text{obs}}^* = 210.9 \cdot 18/47 = 80.77$ Ångström.

11.21 Ur observationerna x_1, \dots, x_{n_1} fås att

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i = \frac{1}{7} \cdot 42 = 6$$

och ur y_1, \dots, y_{n_2} att

$$\bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i = \frac{1}{4} \cdot 8 = 2.$$

Vidare så är

$$s^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

en väntevärdesriktig skattning av σ^2 . Den första summan beräknas genom iakttagelsen att

$$\sum_{i=1}^{n_1} (x_i - \bar{x})^2 = \sum_{i=1}^{n_1} (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^{n_1} x_i^2 - n_1\bar{x}^2 = 276 - 7 \cdot 6^2 = 24.$$

På samma sätt är

$$\sum_{i=1}^{n_2} (y_i - \bar{y})^2 = \sum_{i=1}^{n_2} y_i^2 - n_2\bar{y}^2 = 19 - 4 \cdot 2^2 = 3$$

så

$$s^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{24 + 3}{6 + 4 - 2} = \frac{27}{8}$$

och $s = \sqrt{27/8} = 1.8371$ är en skattning av σ .

11.22 Låt $y_i^{(j)}$ vara mätresultat $i, i = 1, \dots, n$, i mätserie $j, j = 1, 2, 3$. Dessa modelleras som utfall av oberoende stokastiska variabler $Y_i^{(j)}$ där $Y_i^{(j)}$ är $N(\mu_j, \sigma)$. Väntevärdena μ_1, μ_2, μ_3 skattas med

$$\begin{aligned} \bar{y}^{(1)} &= \frac{1}{n} \sum_{i=1}^n y_i^{(1)} = 0.184 \\ \bar{y}^{(2)} &= \frac{1}{n} \sum_{i=1}^n y_i^{(2)} = 0.232 \\ \bar{y}^{(3)} &= \frac{1}{n} \sum_{i=1}^n y_i^{(3)} = 0.158 \end{aligned}$$

respektive. De tre mäteserierna ger tre skattningar av σ :

$$\begin{aligned} s_1 &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i^{(1)} - \bar{y}^{(1)})^2} = 0.018166 \\ s_2 &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i^{(2)} - \bar{y}^{(2)})^2} = 0.016432 \\ s_3 &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i^{(3)} - \bar{y}^{(3)})^2} = 0.016432 \end{aligned}$$

Variansskattningarna s_1^2, s_2^2 och s_3^2 poolas samman till en skattning av σ^2

$$s^2 = \frac{(n-1)s_1^2 + (n-1)s_2^2 + (n-1)s_3^2}{(n-1) + (n-1) + (n-1)} = \frac{s_1^2 + s_2^2 + s_3^2}{3} = 0.00029,$$

det vill säga $s = 0.017029$ är skattningen av σ .

11.23 Låt $x = 16$ vara ett utfall av en binomialfördelad stokastisk variabel X, X är $\text{Bin}(n, p)$ där $n = 25$.

ML-skattning: Den logaritmerade likelihoodfunktionen är

$$\ln(L(p)) = \ln \binom{n}{x} + x \ln(p) + (n-x) \ln(1-p)$$

så lösning av

$$0 = \frac{d}{dp} \ln(L(p)) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{1}{p(1-p)} [x - np]$$

ger, då $p(1-p) \neq 0$, att $p = x/n$. Alltså, maximum-likelihoodskattningen är $p_{\text{obs}}^* = x/n$.

MK-skattning: Väntevärdet $E(X) = np$ så minsta-kvadratskattningen av p är det värde på p som minimerar

$$Q(p) = (x - E(X))^2 = (x - np)^2.$$

Eftersom $(x - np)^2 \geq 0$ med likhet om $x = np$ fås skattningen $p_{\text{obs}}^* = x/n$.

Med värden fås skattningen $p_{\text{obs}}^* = x/n = 16/25 = 0.64$

b) Skattningen $p_{\text{obs}}^* = x/n$ är ett utfall av en stokastisk variabel med varians

$$V(p^*) = V\left(\frac{1}{n}X\right) = \frac{1}{n^2} \underbrace{V(X)}_{=np(1-p)} = \frac{p(1-p)}{n}$$

så standardavvikelsen är $D(p^*) = \sqrt{p(1-p)/n}$.

c) En skattning av $D(p^*) = \sqrt{p(1-p)/n}$ är

$$\sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n}} = \sqrt{\frac{0.64(1-0.64)}{25}} = 0.096$$

och kallas medelfelet för skattningen p_{obs}^* .

11.24 Låt x_1 och x_2 vara antalet färgblinda som forskare A respektive B har funnit. Vi ansätter modellen att x_1 och x_2 är utfall av oberoende stokastiska variabler X_1 och X_2 där X_1 är $\text{Bin}(n_1, p) = \text{Bin}(1000, p)$ och X_2 är $\text{Bin}(n_2, p) = \text{Bin}(2000, p)$.

Likelihoodfunktionen är

$$L(p) = p_{X_1, X_2}(x_1, x_2) = \{\text{oberoende}\} = p_{X_1}(x_1)p_{X_2}(x_2) = \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \binom{n_2}{x_2} p^{x_2} (1-p)^{n_2-x_2}$$

så den logaritmerade likelihoodfunktionen är

$$\ln(L(p)) = \ln \left(\binom{n_1}{x_1} \binom{n_2}{x_2} \right) + (x_1 + x_2) \ln(p) + (n_1 + n_2 - x_1 - x_2) \ln(1-p).$$

Lösning av

$$0 = \frac{d}{dp} \ln(L(p)) = \frac{x_1 + x_2}{p} - \frac{(n_1 + n_2) - (x_1 + x_2)}{1-p} = \frac{1}{p(1-p)} [(x_1 + x_2) - (n_1 + n_2)p]$$

ger skattningen $p_{\text{obs}}^* = (x_1 + x_2)/(n_1 + n_2) = 201/3000 = 0.067$. Notera att detta kan ses som att man slår samman de två observationsserierna till en och betraktar den relativa frekvensen av färgblinda i den sammanslagna serien.

11.25 Låt x_1, \dots, x_n vara antalet fartyg som passerat Helsingborg under tidsperioder av längder t_1, \dots, t_n . Vi ansätter modellen att x_1, \dots, x_n är utfall av oberoende Poissonfördelade stokastiska variabler X_1, \dots, X_n där $E(X_i) = \lambda t_i$.

Likelihoodfunktionen blir

$$L(\lambda) = p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \{\text{oberoende}\} = p_{X_1}(x_1) \cdots p_{X_n}(x_n) = \frac{(\lambda t_1)^{x_1}}{x_1!} e^{-\lambda t_1} \cdots \frac{(\lambda t_n)^{x_n}}{x_n!} e^{-\lambda t_n}.$$

Det λ som maximerar $L(\lambda)$ är samma λ som maximerar

$$\ln(L(\lambda)) = \sum_{i=1}^n (x_i \ln(\lambda) + x_i \ln(t_i) - \ln(x_i!) - \lambda t_i).$$

Detta maximum bestäms som det λ som löser

$$0 = \frac{d}{d\lambda} \ln(L(\lambda)) = \sum_{i=1}^n \left(\frac{x_i}{\lambda} - t_i \right) = \frac{1}{\lambda} \left(\sum_{i=1}^n x_i \right) - \left(\sum_{i=1}^n t_i \right)$$

vilket ger skattningen

$$\lambda_{\text{obs}}^* = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n t_i}.$$

Med de $n = 3$ observationerna

Observationstid, t_i	30	30	40
Antal fartyg x_i	10	12	18

fås skattningen $\lambda_{\text{obs}}^* = 40/100 = 0.40$ fartyg per minut.

b) Skattningen λ_{obs}^* är ett utfall av en stokastisk variabel med varians

$$V(\lambda^*) = V\left(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n t_i}\right) = \{\text{oberoende}\} = \frac{1}{\left(\sum_{i=1}^n t_i\right)^2} \sum_{i=1}^n \underbrace{V(X_i)}_{=\lambda t_i} = \frac{\lambda}{\sum_{i=1}^n t_i},$$

det vill säga med standardavvikelse

$$D(\lambda^*) = \sqrt{\frac{\lambda}{\sum_{i=1}^n t_i}} = \frac{\sqrt{\lambda}}{10}.$$

11.26 Låt $x = 32$ vara antalet bostadssökande av $n = 50$ undersökta som redan har en bostad. Ansätt modellen att x är ett utfall av en hypergeometriskt fördelad stokastisk variabel X . Låt N beteckna antalet personer i bostadskö och p andelen av dessa som redan har en bostad.

Väntevärdet $E(X) = np$ så minsta-kvadratskattningen av p är det värde på p som minimerar

$$Q(p) = (x - E(X))^2 = (x - np)^2.$$

Eftersom $(x - np)^2 \geq 0$ med likhet om $x = np$ fås skattningen $p_{\text{obs}}^* = x/n$.

a) Med värden fås skattningen $p_{\text{obs}}^* = x/n = 32/50 = 0.64$

b) Eftersom $V(X) = np(1-p)(N-n)/(N-1)$ är skattningen p_{obs}^* är ett utfall av en stokastisk variabel med varians

$$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{p(1-p)}{n} \frac{N-n}{N-1}.$$

Standardavvikelsen $D(p^*) = \sqrt{\frac{p(1-p)}{n} \frac{N-n}{N-1}}$ skattas då $N = 100$ med

$$\sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n} \frac{N-n}{N-1}} = \sqrt{\frac{0.64(1-0.64)}{50} \frac{100-50}{100-1}} = 0.048242$$

c) och då $N = 1000$ med

$$\sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n} \frac{N-n}{N-1}} = \sqrt{\frac{0.64(1-0.64)}{50} \frac{1000-50}{1000-1}} = 0.066197.$$

11.27 Smältpunktsmätningarna x_1, \dots, x_m modelleras som utfall av oberoende och likafördelade stokastiska variabler X_1, \dots, X_n med väntevärde $\mu = E(X_i)$ och standardavvikelse $\sigma = D(X_i)$. Väntevärdet skattas med $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 80.9$ som är ett utfall av $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ där

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}$$

så $D(\bar{X}) = \sigma/\sqrt{n}$. Denna standardavvikelse skattas med $d(\bar{X}) = s/\sqrt{n} = 0.3/\sqrt{9} = 0.1$ och kallas för medelfelet för \bar{x} .

11.28 Mätningarna av konservburkarnas vikter, x_1, \dots, x_n , modelleras som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n .

Maximum-likelihoodskattningarna av μ och σ är de parametervärden som maximerar funktionen

$$\begin{aligned} L(\mu, \sigma) &= f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \{\text{oberoende}\} = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2} = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}. \end{aligned}$$

Lösning av

$$0 = \frac{\partial}{\partial \mu} \ln(L(\mu, \sigma)) = \frac{\partial}{\partial \mu} \left[-n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{n}{\sigma^2} [\bar{x} - \mu]$$

ger $\mu = \bar{x}$ då $\sigma > 0$. Alltså, ML-skattningen av μ är $\bar{x} = 250.32$. Lösning av

$$\begin{aligned} 0 &= \frac{\partial}{\partial \sigma} \ln(L(\mu, \sigma)) = \frac{\partial}{\partial \sigma} \left[-n \ln(\sqrt{2\pi}) - n \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 \\ &= \frac{-n}{\sigma^3} \left[\sigma^2 - \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right] \end{aligned}$$

vilket ger

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

som tillsammans med skattningen av μ ger ML-skattningen av σ är

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = 2.1106.$$

11.29 Låt x_1, \dots, x_n vara resultaten av mätningarna av kvadratens sida. Dessa antas vara utfall av oberoende $N(\sqrt{\theta}, \sigma)$ -fördelade stokastiska variabler där σ är känd och θ är kvadratens (okända) area.

Maximum-likelihoodskattningen av θ är det θ som maximerar

$$\begin{aligned} L(\theta) &= f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \{\text{oberoende}\} = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_1 - \sqrt{\theta})^2 / 2\sigma^2} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_n - \sqrt{\theta})^2 / 2\sigma^2}. \end{aligned}$$

Det är samma θ som maximerar

$$\ln(L(\theta)) = \ln\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i - \sqrt{\theta})^2/2\sigma^2}\right) = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \sqrt{\theta})^2$$

och detta maximum bestäms som lösningen till

$$0 = \frac{d}{d\theta} \ln(L(\theta)) = \frac{1}{2\sqrt{\theta}\sigma^2} \sum_{i=1}^n (x_i - \sqrt{\theta}) = \frac{1}{2\sqrt{\theta}\sigma^2} \left(\underbrace{\sum_{i=1}^n x_i}_{=n\bar{x}} - n\sqrt{\theta} \right) = \frac{n}{2\sqrt{\theta}\sigma^2} (\bar{x} - \sqrt{\theta}),$$

vilket ger $\sqrt{\theta} = \bar{x}$ eller $\theta = \bar{x}^2$. Alltså, ML-skattningen av θ är $\theta_{\text{obs}}^* = (\bar{x})^2$.

b) Skattningen θ_{obs}^* är inte väntevärdesriktig eftersom

$$E(\theta^*) = E(\bar{X}^2) = V(\bar{X}) + (E(\bar{X}))^2 = \frac{\sigma^2}{n} + (\sqrt{\theta})^2 = \frac{\sigma^2}{n} + \theta \neq \theta,$$

men med $\hat{\theta}_{\text{obs}} = \bar{x}^2 - \sigma^2/n$ fås en väntevärdesriktig skattning.

11.30 Låt X beskriva antalet dragna kulor med olika färg tills två med samma färg erhålls. Möjliga värden på X ges av $\Omega_X = \{1, 2, \dots, N\}$. (Notera att antalet dragna kulor är $X + 1$.)

Händelsen $X = 1$ är att de två första kulorna har samma färg, det vill säga $P(X = 1) = 1/N$. Händelsen $X = k$, $k > 1$ är att de första k kulorna har olika färg och kula $k + 1$ har någon av de tidigare k dragna färgerna. Alltså,

$$P(X = k) = \underbrace{\frac{N}{N} \cdot \frac{N-1}{N} \cdots \frac{N-(k-1)}{N}}_{k \text{ st}} \cdot \frac{k}{N} = \frac{N!}{(N-k)!N^k} \cdot \frac{k}{N}, \quad k = 1, 2, \dots, N.$$

Med utfallet $x = 3$ blir likelihoodfunktionen

$$L(N) = \frac{N!}{(N-x)!N^x} \cdot \frac{x}{N} = \frac{(N-1)(N-2)3}{N^3}$$

vilken är avtagande för stora N . För olika värden på N fås

N	$L(N)$
3	0.22222
4	0.28125
5	0.28800
6	0.27778
7	0.26239
8	0.24609

Maximum ges med valet $N = 5$.

Anmärkning. Om man vill få en uppfattning om var detta maximum ligger någonstans kan man för tillfället låta N vara kontinuerlig i uttrycket $\frac{(N-1)(N-2)3}{N^3}$. Logaritmering och derivering ger att maximum nås i det N som löser

$$0 = \frac{d}{dN} \ln(L(N)) = \frac{d}{dN} [\ln(N-1) + \ln(N-2) + \ln(3) - 3 \ln(N)] = \frac{-(N-3-\sqrt{3})(N-3+\sqrt{3})}{N(N-1)(N-2)}$$

det vill säga $N = 3 + \sqrt{3} = 4.73$. Alltså erhålls tillåtet maximum då $N = 4$ eller $N = 5$. Av dessa två är $L(N)$ störst då $N = 5$.

11.31 De $n = 10$ observationerna $(x_1, y_1), \dots, (x_n, y_n)$ modelleras som utfall av oberoende

$$N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_{xx}^2 & \sigma_{xy}^2 \\ \sigma_{yx}^2 & \sigma_{yy}^2 \end{bmatrix}\right)\text{-fördelade}$$

stokastiska variabler $(X_1, Y_1), \dots, (X_n, Y_n)$. Väntevärdet $\mu_x = E(X_i)$ skattas med $\bar{x} = 4.899$ och variansen $\sigma_{xx}^2 = V(X_i)$ skattas med $s_x^2 = 2.8254$, det vill säga $D(X_i)$ skattas med $s_x = 1.6809$.

Väntevärdet $\mu_y = E(Y_i)$ skattas med $\bar{y} = 13.617$ och variansen $\sigma_{yy}^2 = V(Y_i)$ skattas med $s_y^2 = 17.674$, det vill säga $D(Y_i)$ skattas med $s_y = 4.2041$.

Kovariansen $\sigma_{xy}^2 = C(X_i, Y_i)$ skattas med

$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) = 3.0966$$

så korrelationen $\rho = \frac{\sigma_{xy}^2}{\sigma_{xx} \sigma_{yy}}$ skattas med

$$r = \frac{c_{xy}}{s_x s_y} = \frac{3.0966}{1.6809 \cdot 4.2104} = 0.4382.$$

Givet initieringen

```
r = 0.4382;
mx = 4.899; sx = 1.6809;
my = 13.617; sy = 4.2041;
korr = [];
```

kan följande kodsnutt utföras 10000 gånger

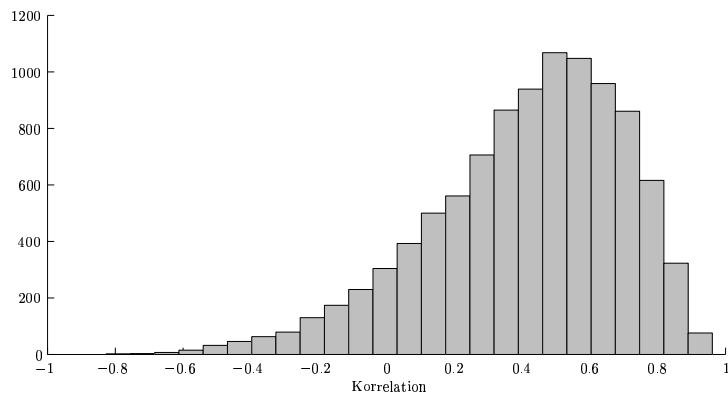
```
n=10;
z1 = randn(1,n);
x = sx*z1+mx;
z2 = randn(1,n);
y = sy*(r*z1 + sqrt(1-r^2)*z2) + my;

korr = [korr (x-mean(x))*(y-mean(y))' / ...
        sqrt((x-mean(x))*(x-mean(x))'*(y-mean(y))*(y-mean(y))')];
```

Detta ger 10000 skattningar av korrelationen för stickprov om storlek 10 från den (empiriska) tvådimensionella normalfördelningen.

Dessa skattningar har medelfel

```
>> std(korr)
ans =
    0.28731
```

Ett histogram av 10000 korrelationskattningar för stickprov av storlek 10.

12.1 Låt X vara χ^2 -fördelad med 24 frihetsgrader. Bestäm a så att $P(X < a) = 0.95$. Då är

$$0.05 = 1 - P(X \leq a) = 1 - F_X(a) = 1 - F_X(\chi_{0.05}^2)$$

det vill säga $a = \chi_{0.05}^2 = 36.415$.

Bestäm b och c , $b < c$, så att $P(b < X < c) = 0.95$. Vi antar att händelsen $X \notin (b, c)$ är sådan att $P(X \leq b) = 0.025$ och $P(X \geq c) = 0.025$.

Då är

$$0.025 = 1 - P(X \leq c) = 1 - F_X(c) = 1 - F_X(\chi_{0.025}^2)$$

dvs $c = \chi_{0.025}^2 = 39.364$ och

$$0.975 = 1 - P(X \leq b) = 1 - F_X(b) = 1 - F_X(\chi_{0.975}^2)$$

dvs $b = \chi_{0.975}^2 = 12.401$.

12.2 Om Z, Z_1, Z_2, \dots, Z_n är oberoende $N(0, 1)$ så är

$$X = \sum_{i=1}^n Z_i^2$$

χ^2 -fördelad med n frihetsgrader, $X \sim \chi^2(n)$. För Z är

$$E(Z^2) = V(Z) + (E(Z))^2 = 1 + 0^2 = 1$$

så

$$E(X) = E\left(\sum_{i=1}^n Z_i^2\right) = \sum_{i=1}^n \underbrace{E(Z_i^2)}_{=1} = n.$$

Vidare, enligt ledning är $E(Z^4) = 3$ vilket medför att

$$V(Z^2) = E(Z^4) - (E(Z^2))^2 = 3 - 1^2 = 2$$

och

$$V(X) = V\left(\sum_{i=1}^n Z_i^2\right) = \{\text{oberoende}\} = \sum_{i=1}^n \underbrace{V(Z_i^2)}_{=2} = 2n.$$

12.3 Om Z, Z_1, Z_2, \dots, Z_n är oberoende $N(0, 1)$ så är

$$X = \sum_{i=1}^n Z_i^2$$

$\chi^2(n)$ -fördelad. Ur uppgift 12.2 är

$$E(X) = n \quad V(X) = 2n,$$

och enligt centrala gränsvärdesatsen är X approximativt normalfördelad, X är approximativt $N(n, \sqrt{2n})$. Det vill säga

$$1 - \Phi(\lambda_\alpha) = \alpha = P(X > \chi_\alpha^2) = P\left(\frac{X - n}{\sqrt{2n}} > \frac{\chi_\alpha^2 - n}{\sqrt{2n}}\right) \approx 1 - \Phi\left(\frac{\chi_\alpha^2 - n}{\sqrt{2n}}\right)$$

ger att

$$\lambda_\alpha \approx \frac{\chi_\alpha^2 - n}{\sqrt{2n}}$$

eller

$$\chi_\alpha^2 \approx n + \sqrt{2n}\lambda_\alpha.$$

12.4 Låt X vara $t(9)$ -fördelad. Bestäm a så att $P(|X| \leq a) = 0.99$. Då är

$$\begin{aligned} 0.99 &= P(|X| \leq a) = P(-a \leq X \leq a) = F_X(a) - F_X(-a) = F_X(a) - (1 - F_X(a)) \\ &= 2F_X(a) - 1. \end{aligned}$$

Alltså är $F_X(a) = 0.995$ och

$$0.005 = 1 - F_X(a) = 1 - F_X(t_{0.005})$$

dvs $a = t_{0.005} = 3.2498$.

Med

$$0.05 = P(X > b) = 1 - F_X(b) = 1 - F_X(t_{0.05})$$

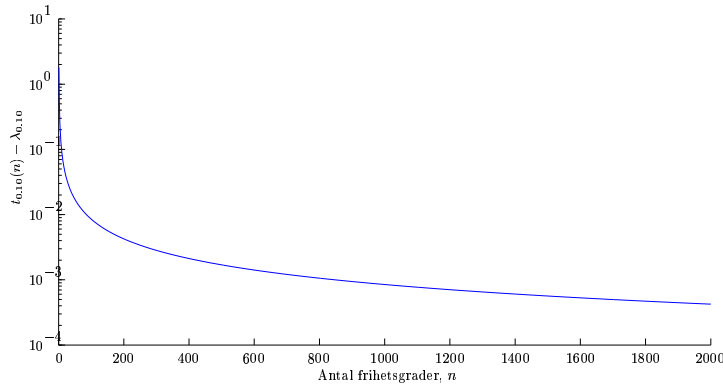
fås $b = t_{0.05} = 1.8331$.

Med tabeller och korrekta avrundningsregler fås följande

Kvantil $\alpha = 0.025$					
# decimaler	$\lambda_\alpha = t_\alpha$	# frihetsgrader	# decimaler	$\lambda_\alpha = t_\alpha$	# frihetsgrader
1	2.0	28	1	1.3	14
2	1.96	473	2	1.28	247
3	1.960	4427	3	1.282	894
4	1.9600	27581	4	1.2816	8602

Kvantil $\alpha = 0.05$			Kvantil $\alpha = 0.10$		
# decimaler	$\lambda_\alpha = t_\alpha$	# frihetsgrader	# decimaler	$\lambda_\alpha = t_\alpha$	# frihetsgrader
1	1.6	298	1	1.3	14
2	1.64	10412	2	1.28	247
3	1.645	10412	3	1.282	894
4	1.6449	15813	4	1.2816	8602

En bättre bild över hur $t(n)$ -fördelningen närmar sig normalfördelningen då n växer fås av att betrakta skillnaden $t_\alpha(n) - \lambda_\alpha$ som tal inte i antalet korrekta decimaler. Nedan visas skillnaden för växande antal frihetsgrader.



Skillnaden $t_{0.10}(n) - \lambda_{0.10}$ för växande antal frihetsgrader.

12.5 Låt x beteckna antalet konfidensintervall som innehåller den därför avsedda konstanten. Då är x ett utfall av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(n, p)$, där $n = 15$ är antalet konfidensintervall och $p = 0.90$ är konfidensgraden för varje intervall.

a) Med X som $\text{Bin}(n, p) = \text{Bin}(15, 0.90)$ är $P(X = n) = p^n = 0.90^{15} = 0.20589$.

b) Typvärdet k , det värde som maximerar $P(X = k)$, är $k = 14$ då $P(X = 14) = 15p^{14}(1-p) = 0.34315$, alltså 14 av 15 intervall träffar och ett intervall missar.

12.6 Låt glödlampornas livslängder beskrivas av den stokastiska variabeln X med täthetsfunktion

$$f_X(x) = \frac{1}{\theta} e^{-x/\theta}, x \geq 0.$$

Detta är exponentialfördelningen med intensitet $1/\theta$, det vill säga väntevärde θ . Fördelningsfunktionen är för $t \geq 0$

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \int_0^t \frac{1}{\theta} e^{-x/\theta} dx = 1 - e^{-t/\theta}.$$

Nu är för $c > 0$

$$P(\theta < cX) = P\left(X > \frac{\theta}{c}\right) = 1 - F_X\left(\frac{\theta}{c}\right) = e^{-(\theta/c)/\theta} = e^{-1/c}.$$

Alltså, med $c = -1/\ln(0.975)$ är $P(\theta < cX) = 0.975$ och med $c = -1/\ln(0.025)$ är $P(\theta < cX) = 0.025$. Alltså,

$$P\left(\frac{X}{-\ln(0.025)} < \theta < \frac{X}{-\ln(0.975)}\right) = 0.95$$

och ett 95% konfidensintervall för θ är

$$\frac{x}{-\ln(0.025)} < \theta < \frac{x}{-\ln(0.975)}.$$

Med observationen $x = 1000$ erhålls intervallet

$$271.09 \leq \theta \leq 39498 \quad (95\%).$$

12.7 Konfidensgraden väljes fortfarande till 95%. Med $c = -1/\ln(0.05)$ är $P(\theta < cX) = 0.05$ vilket ger

$$P\left(\frac{X}{-\ln(0.05)} < \theta\right) = 0.95$$

och ett 95% konfidensintervall för θ är

$$\frac{x}{-\ln(0.05)} < \theta.$$

Med observationen $x = 1000$ erhålls intervallet

$$333.81 \leq \theta \quad (95\%)$$

eller, omformulerat, $\theta \in [333.81, \infty)$.

12.8 1. Ja. Enligt stora talens lag gäller att andelen intervall som täcker θ går mot $P(\theta \in I_\theta(X_1, \dots, X_n)) = 1 - \alpha$ då antalet intervall ökar.

2. Nej. Intervallet $I_\theta(x_1, \dots, x_n)$ är konstruerat så att $P(\theta \in I_\theta(X_1, \dots, X_n))$ med sannolikhet $1 - \alpha$, inte $P(X_i \in I_\theta(X_1, \dots, X_n))$.

3. Nej. Dels av samma anledning som ovan, dels för att observationerna är utfall av stokastiska variabler och alltså inte säkert hamnar på ett intervall eller inte.

4. Ja. Detta är definitionen av konfidensgraden, $P(\theta \in I_\theta(X_1, \dots, X_n)) = 1 - \alpha$.

12.9 Låt X beskriva resultatet av en avståndsmätning. Modell:

$$X = \text{avstånd} + \text{mätfel} = \mu + \epsilon$$

där ϵ är $N(0, \sigma)$, $\sigma = 5 \cdot 10^{-3}$. Då är $X \sim N(\mu, \sigma)$. Låt X_1, \dots, X_n beskriva resultaten av n oberoende avståndsmätningar med observerade värden x_1, \dots, x_n . Väntevärdet $\mu = E(X)$ skattas med \bar{x} som beskrivs av \bar{X} , där

$$\bar{X} \text{ är } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

så med sannolikhet $1 - \alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1 - \alpha$ är

$$\bar{X} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\bar{x} = 1132.155$. Konfidensgrad $1 - \alpha = 0.95$ ger $\lambda_{\alpha/2} = \lambda_{0.025} = 1.9600$ och intervallet

$$\mu \in \bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1132.155 \pm 1.96 \frac{5 \cdot 10^{-3}}{\sqrt{4}} = 1132.155 \pm 0.0048999 \quad (95\%)$$

eller

$$1132.150 \leq \mu \leq 1132.160 \quad (95\%).$$

12.10 Låt x_1, \dots, x_n vara utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n där $\sigma = 2$. Väntevärdet $\mu = E(X)$ skattas med \bar{x} som beskrivs av \bar{X} , där

$$\bar{X} \text{ är } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

så med sannolikhet $1 - \alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1 - \alpha$ är

$$\bar{X} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\bar{x} = 45.2$. Konfidensgrad $1 - \alpha = 0.95$ ger $\lambda_{\alpha/2} = \lambda_{0.025} = 1.9600$ och intervallet

$$\mu \in \bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 45.2 \pm 1.96 \frac{2}{\sqrt{4}} = 45.2 \pm 1.96 \quad (95\%)$$

eller

$$43.24 \leq \mu \leq 47.16 \quad (95\%).$$

12.11 Låt x_1, \dots, x_n vara utfall av oberoende $N(\mu + \Delta, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n där $\sigma = 0.05$ och $\Delta = 0.10$. Om väntevärdet $\mu + \Delta$ skattas med \bar{x} kan pH-halten μ skattas med $\bar{x} - \Delta$. Denna skattning beskrivs av $\bar{X} - \Delta$, där

$$\bar{X} - \Delta \text{ är } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{(\bar{X} - \Delta) - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

så med sannolikhet $1 - \alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{(\bar{X} - \Delta) - \mu}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1 - \alpha$ är

$$(\bar{X} - \Delta) - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq (\bar{X} - \Delta) + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\bar{x} = 8.2$. Konfidensgrad $1 - \alpha = 0.99$ ger $\lambda_{\alpha/2} = \lambda_{0.005} = 2.5758$ och intervallet

$$\mu \in (\bar{x} - \Delta) \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (8.2 - 0.1) \pm 2.5758 \frac{0.05}{\sqrt{4}} = 8.1 \pm 0.0644 \quad (99\%)$$

eller

$$8.0356 \leq \mu \leq 8.1644 \quad (99\%).$$

12.12 Konfidensintervallet för väntevärdet med konfidensgrad $1 - \alpha$ ges av

$$\bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

det vill säga har bredden

$$L_{1-\alpha, n} = 2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Med konfidensgrad $1 - \alpha = 0.90$ fås $\lambda_{\alpha/2} = \lambda_{0.05} = 1.6449$.

a) Vi söker n så att $L(0.90, n) = L(0.90, 5)/2$ [alternativt $L(0.90, 5)/10$].

$$2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{5}} / 2$$

ger $\sqrt{n} = 2\sqrt{5}$, eller $n = 4 \cdot 5 = 20$. Med en tiondel så brett fås $\sqrt{n} = 10\sqrt{5}$ eller $n = 100 \cdot 5 = 500$.

b) Vi söker n så att $L(0.99, n) = L(0.90, 5)$.

$$2\lambda_{0.005} \frac{\sigma}{\sqrt{n}} = 2\lambda_{0.05} \frac{\sigma}{\sqrt{5}}$$

ger $\sqrt{n} = \sqrt{5}\lambda_{0.005}/\lambda_{0.05}$ eller, med $\lambda_{0.005} = 2.5758$, $n = 5(\lambda_{0.005}/\lambda_{0.05})^2 = 12.26$, det vill säga 13 observationer.

c) Vi söker n så att $L(0.99, n) = L(0.90, 5)/2$ [alternativt $L(0.90, 5)/10$].

$$2\lambda_{0.005} \frac{\sigma}{\sqrt{n}} = 2\lambda_{0.05} \frac{\sigma}{\sqrt{5}} / 2$$

ger $\sqrt{n} = 2\sqrt{5}\lambda_{0.005}/\lambda_{0.05}$ eller, med $\lambda_{0.005} = 2.5758$, $n = 4 \cdot 5(\lambda_{0.005}/\lambda_{0.05})^2 = 49.047$, dvs. 50 stycken. Med en tiondel så brett fås $n = (10)^2 \cdot 5(\lambda_{0.005}/\lambda_{0.05})^2 = 1226.2$, dvs. 1227 stycken.

Notera att i denna lösning utnyttjade vi aldrig det uppmätta intervallet [7.02, 7.14] eftersom vi inte behövde bestämma \bar{x} eller σ .

12.13 Låt x_1, \dots, x_n vara avkastningen i ton under $n = 10$ dagar. Dessa modelleras som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n . Alltså är $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ och

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ är } t(n-1)\text{-fördelad}$$

där S^2 ges av

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Således, med sannolikhet $1 - \alpha$ är

$$-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2},$$

vilket omskrivet ger att med sannolikhet $1 - \alpha$ är

$$\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}.$$

Med $n = 10$ observationer och konfidensgrad $1 - \alpha = 0.95$ fås ur $t(n-1) = t(9)$ -tabell att $t_{\alpha/2} = t_{0.025} = 2.26$. Vidare beräknas

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 7.51 \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.1543$$

$s = \sqrt{s^2} = 0.3929$, så konfidensintervallet blir

$$\mu \in \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 7.51 \pm 0.281 = [7.229, 7.791] \quad (95\%).$$

12.14 Från uppgift 12.11 vet vi att

$$\frac{(\bar{X} - \Delta) - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Standardavvikelsen σ skattas med

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0.042426$$

så

$$\frac{(\bar{X} - \Delta) - \mu}{S/\sqrt{n}} \sim t(n-1)\text{-fördelad}.$$

Med sannolikhet $1 - \alpha$ är

$$(\bar{X} - \Delta) - t_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq (\bar{X} - \Delta) + t_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

där kvantilen $t_{\alpha/2}$ fås ur $t(n-1)$ -tabell.

Med $n = 4$ observationer fås $\bar{x} = 8.2$. Konfidensgrad $1 - \alpha = 0.99$ ger $t_{\alpha/2} = t_{0.005} = 5.8409$ och intervallet

$$\mu \in (\bar{x} - \Delta) \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = (8.2 - 0.1) \pm 5.8409 \frac{0.042426}{\sqrt{4}} = 8.1 \pm 0.1239 \quad (99\%)$$

eller

$$7.9761 \leq \mu \leq 8.2239 \quad (99\%).$$

12.15 MatLab-koden

```
n = 10;
mu = 12;
sigma = 2;
x = randn(n,1)*sigma + mu;
```

skapar $n = 10$ observationer x_1, \dots, x_n från $N(\mu, \sigma)$ -fördelningen. Ett 95% konfidensintervall för μ ges av

$$\mu \in \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 12.618 \pm 1.96 \cdot \frac{2}{\sqrt{10}} = \bar{x} \pm 1.96 \cdot \frac{2}{\sqrt{10}} = 12.618 \pm 1.2396 = (11.378, 13.857).$$

I detta fall täcker intervallet värdet $\mu = 12$.

När detta upprepades 1000 gånger kom i 962 av fallen intervallet att innehålla $\mu = 12$ och i 38 av fallen det inte.

Med 2.5%-kvantilen 1.96 i normalfördelningen befanns 928 av de 1000 intervallen vara sådana att

$$12 \in \bar{x} \pm 1.96 \cdot \frac{s}{\sqrt{10}}$$

och 953 av samma 1000 intervall uppfyllde

$$12 \in \bar{x} \pm t_{0.025} \cdot \frac{s}{\sqrt{10}}$$

där $t_{0.025}$ bestämdes ur $t(n-1) = t(9)$ -tabell till $t_{0.025} = 2.2622$.

12.16 Baserat på observationer x_1, \dots, x_n av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler ges ett konfidensintervall, med konfidensgrad $1 - \alpha$, för μ av

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}.$$

Intervallet $(\bar{x} - s, \bar{x} + s)$ har konfidensgrad åtminstone 99% om n är sådan att $t_{0.005}/\sqrt{n} < 1$.

n	$t_{0.005}$	$t_{0.005}/\sqrt{n}$	n	$t_{0.005}$	$t_{0.005}/\sqrt{n}$
2	63.657	45.012	8	3.4995	1.2373
3	9.9248	5.7301	9	3.3554	1.1185
4	5.8409	2.9205	10	3.2498	1.0277
5	4.6041	2.0590	11	3.1693	0.9556
6	4.0321	1.6461	12	3.1058	0.8966
7	3.7074	1.4013	13	3.0545	0.8472

Ur tabellen får vi att $n \geq 11$.

12.17 Skattningen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

av σ^2 beskrivs av S^2 som är sådan att $(n-1)S^2/\sigma^2$ är en $\chi^2(n-1)$ -fördelad stokastisk variabel. Alltså kan man ur $\chi^2(n-1)$ -tabeller bestämma kvantilerna $\chi_{1-\alpha/2}^2$ och $\chi_{\alpha/2}^2$ så att

$$P\left(\chi_{1-\alpha/2}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{\alpha/2}^2\right) = 1 - \alpha$$

eller, omformulerat, med sannolikhet $1 - \alpha$ är

$$\frac{(n-1)S^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}.$$

Med $n = 8$ observationer och skattningen $s = 5.2$ fås ur $\chi^2(7)$ -tabell att $\chi_{0.975}^2 = 1.69$ och $\chi_{0.025}^2 = 16.0$. Alltså är ett 95% konfidensintervall för σ^2

$$11.821 = \frac{(8-1)(5.2)^2}{16.0} \leq \sigma^2 \leq \frac{(8-1)(5.2)^2}{1.69} = 112.01.$$

Motsvarande intervall för σ är

$$3.44 = \sqrt{11.821} \leq \sigma \leq \sqrt{112.01} = 10.6 \quad (95\%).$$

12.18 Låt x_1, \dots, x_n vara furuplankornas uppmätta längder. Dessa modelleras som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n . Väntevärdet μ skattas med \bar{x} som beskrivs av \bar{X} som är $N(\mu, \sigma/\sqrt{n})$. Alltså är

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \quad t(n-1)\text{-fördelad.}$$

och med sannolikhet $1 - \alpha$ är

$$-t_{\alpha/2}(n-1) \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2}(n-1)$$

eller, omformulerat, med sannolikhet $1 - \alpha$ är

$$\mu \in \bar{X} \pm t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}.$$

Med observationer x_1, \dots, x_n , $n = 16$:

$$\begin{array}{cccccccc} 5.8 & 5.9 & 5.1 & 3.5 & 4.2 & 4.9 & 5.3 & 5.3 \\ 4.7 & 3.9 & 4.5 & 4.1 & 4.0 & 4.2 & 4.7 & 4.8 \end{array}$$

fås skattningarna

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 4.6812$$

och

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{0.46962} = 0.68529.$$

Med konfidensgrad $1 - \alpha = 0.95$ fås att $t_{\alpha/2}(15) = t_{0.025}(15) = 2.1314$ och konfidensintervallet blir

$$\mu \in 4.6812 \pm 2.1314 \frac{0.68529}{\sqrt{16}} = 4.6812 \pm 0.36517 = [4.3161, 5.0464] \quad (95\%).$$

b) Vi vet att

$$\frac{(n-1)S^2}{\sigma^2} \text{ är } \chi^2(n-1)\text{-fördelad}$$

om $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ där X_i är oberoende $N(\mu, \sigma)$. Alltså är

$$P\left(\chi_{1-\alpha/2}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{\alpha/2}^2\right) = 1 - \alpha.$$

Med $\alpha = 0.05$ fås ur $\chi^2(n-1) = \chi^2(15)$ -tabeller

$$P\left(6.26 \leq \frac{(n-1)S^2}{\sigma^2} \leq 27.5\right) = 0.95.$$

Alltså med sannolikhet 95% är

$$\frac{(n-1)S^2}{27.5} \leq \sigma^2 \leq \frac{(n-1)S^2}{6.26}.$$

Med skattningen $s^2 = 0.46962$ fås intervallet

$$0.25627 = \frac{(n-1)s^2}{27.5} \leq \sigma^2 \leq \frac{(n-1)s^2}{6.26} = 1.1249 \quad (95\%)$$

eller

$$0.50623 = \sqrt{0.25627} \leq \sigma \leq \sqrt{1.1249} = 1.0606 \quad (95\%).$$

12.19 Vi vet att $(n-1)S^2/\sigma^2$ är $\chi^2(n-1)$ -fördelad. Det vill säga med sannolikhet $1 - \alpha$ är $(n-1)S^2/\sigma^2 \geq \chi_\alpha^2$ eller, omformulerat, $\sigma^2 \leq (n-1)S^2/\chi_\alpha^2$, alternativt,

$$\sigma \leq \sqrt{\frac{(n-1)S^2}{\chi_\alpha^2}}.$$

För $n = 50$ skruvar och $1 - \alpha = 0.95$ fås ur $\chi^2(n-1) = \chi^2(49)$ -tabell att $\chi_\alpha^2 = \chi_{0.05}^2 = 33.93$. Alltså ges ett 95% ensidigt konfidensintervall för σ av

$$\sigma \leq \sqrt{\frac{(n-1)s^2}{\chi_\alpha^2}} = \sqrt{\frac{(50-1)(0.021)^2}{33.93}} = \frac{7 \cdot 0.021}{\sqrt{33.93}} = 0.025.$$

12.20 Ett konfidensintervall för σ ges av

$$\sqrt{\frac{n-1}{\chi_{\alpha/2}^2}} s \leq \sigma \leq \sqrt{\frac{n-1}{\chi_{1-\alpha/2}^2}} s$$

vilket ger att

$$\frac{n-1}{\chi_{0.025}^2} = 0.9^2 = 0.81 \quad \frac{n-1}{\chi_{1-0.025}^2} = 1.1^2 = 1.21.$$

Enligt anmärkningen (och uppgift 12.3) är

$$\chi_\alpha^2 \approx n + \sqrt{2n}\lambda_\alpha$$

för stora n så vi söker n så att

$$\frac{n-1}{n + \sqrt{2n}\lambda_{0.025}} = 0.81 \quad \text{och} \quad \frac{n-1}{n - \sqrt{2n}\lambda_{0.025}} = 1.21.$$

Den vänstra ekvationen har lösningen $n = 149.98$ den högra lösningen $n = 245.46$.

12.21 Låt X_1, \dots, X_{n_x} beskriva de uppmätta överhöjningarna för betongelement från fabrik A och Y_1, \dots, Y_{n_y} motsvarande för fabrik B. Modell: alla stokastiska variabler är oberoende och X_i är $N(\mu_x, \sigma)$ och Y_i är $N(\mu_y, \sigma)$. Då är

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \quad t(n_x + n_y - 2)\text{-fördelad,}$$

där

$$S^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}.$$

Alltså är med sannolikhet $1 - \alpha$

$$\mu_x - \mu_y \in \bar{X} - \bar{Y} \pm t_{\alpha/2} S \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}.$$

Med observationer

$$\bar{x} = 18.1, s_x = 5.0, n_x = 9 \quad \bar{y} = 14.6, s_y = 7.1, n_y = 16$$

fås

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} = \sqrt{41.572} = 6.4476.$$

Ur $t(n_x + n_y - 2) = t(23)$ -tabeller fås att med konfidensgrad $1 - \alpha = 0.99$ är $t_{\alpha/2} = t_{0.005} = 2.8073$ så konfidensintervallet blir

$$\mu_x - \mu_y \in 18.1 - 14.6 \pm 2.8073 \cdot 6.4476 \sqrt{\frac{1}{9} + \frac{1}{16}} = 3.5 \pm 7.54 \quad (99\%)$$

eller intervallet $[-4.04, 11.0]$.

12.22 Blodtrycksmätningarna före och efter behandling sammanfattas i tabellen:

Person nr, i	1	2	3	4	5	6	7	8	9	10
Blodtryck före, x_i	75	70	75	65	95	70	65	70	65	90
Blodtryck efter, y_i	85	70	80	80	100	90	80	75	90	100
Förändring, $w_i = y_i - x_i$	10	0	5	15	5	20	15	5	25	10

Vi modellerar x_i som utfall av en $N(\mu_i, \sigma_1)$ -fördelad stokastiska variabel X_i , $i = 1, \dots, n$, och y_i som ett utfall av en $N(\mu_i + \Delta, \sigma_2)$ -fördelad stokastiska variabel Y_i , $i = 1, \dots, n$. Parametrarna μ_i är personens blodtryck före behandling och Δ är den genomsnittliga behandlingseffekten. Blodtrycksförändringarna $w_i = y_i - x_i$ är utfall av oberoende $N(\Delta, \sigma)$ -fördelade stokastiska variabler W_1, \dots, W_n , $i = 1, \dots, n$.

Vi skattar Δ med $\Delta_{\text{obs}}^* = \bar{w} = 11$ och σ med

$$s_w = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2} = 7.746.$$

Ur $t(n-1) = t(9)$ -tabell fås att $t_{0.025} = 2.2622$ och ett 95% konfidensintervall för Δ ges av

$$\Delta \in \bar{w} \pm t_{0.025} \frac{s}{\sqrt{n}} = 11 \pm 2.26 \cdot \frac{7.746}{\sqrt{10}} = 11 \pm 5.5 \quad (95\%).$$

12.23 Mätningarna x_1, \dots, x_n på lösningen med okänt pH-värde modelleras av oberoende $N(\mu + \Delta, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n .

Vidare, låt y_1, \dots, y_m vara de sex bestämningarna på en lösning med det kända pH-värdet 4.84. Dessa modelleras som utfall av oberoende $N(4.84 + \Delta, \sigma)$ -fördelade stokastiska variabler Y_1, \dots, Y_m .

En skattning av Δ är $\Delta_{\text{obs}}^* = \bar{y} - 4.84 = 4.7 - 4.84 = -0.14$. Detta är en väntevärdesriktig skattning ty

$$E(\Delta^*) = E(\bar{Y} - 4.84) = E(\bar{Y}) - 4.84 = 4.84 + \Delta - 4.84 = \Delta.$$

Vidare så är

$$V(\Delta^*) = V(\bar{Y} - 4.84) = V(\bar{Y}) = \frac{\sigma^2}{m}.$$

a) En skattning av $\mu + \Delta$ är $\bar{x} = 4.285$ så en skattning av μ fås som $\mu_{\text{obs}}^* = \bar{x} - \Delta_{\text{obs}}^* = 4.425$. Denna är väntevärdesriktig ty

$$E(\mu^*) = E(\bar{X} - \Delta^*) = E(\bar{X}) - E(\Delta^*) = (\mu + \Delta) - \Delta = \mu.$$

b) Här är

$$V(\mu^*) = V(\bar{X} - \Delta^*) = \{\text{oberoende}\} = V(\bar{X}) + (-1)^2 V(\Delta^*) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \frac{(n+m)\sigma^2}{nm}$$

så $D(\mu^*) = \sigma \sqrt{\frac{n+m}{nm}}$.

c) För att skatta $D(\mu^*)$ måste σ skattas. Från x_1, \dots, x_n har vi skattningen

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0.072342$$

och från y_1, \dots, y_m har vi att σ kan skattas med

$$s_y = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2} = 0.089443.$$

Båda dessa skattningar kombineras till s , där

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{(n-1) + (m-1)} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2} = 0.0069625$$

så $s = 0.083442$. Standardavvikelsen $D(\mu^*) = \sigma \sqrt{\frac{n+m}{nm}}$ skattas med

$$s \sqrt{\frac{n+m}{nm}} = 0.083442 \sqrt{\frac{4+6}{24}} = 0.053861,$$

det vill säga medelfelet $d(\mu^*) = 0.054$.

d) Ur $t(n+m-2) = t(8)$ -tabeller fås $t_{0.025} = 2.306$ och ett 95% konfidensintervall för μ ges av

$$\mu \in \mu_{\text{obs}}^* \pm t_{0.025} d(\mu^*) = 4.425 \pm 2.306 \cdot 0.053861 = 4.425 \pm 0.124 \quad (95\%).$$

12.24 Observationerna x_1, \dots, x_n och y_1, \dots, y_n sammanfattas av tabellen:

Förare, i	1	2	3	4	5
Förslitning däcktyp A, x_i	1.0	0.9	0.7	1.5	0.5
Förslitning däcktyp B, y_i	0.9	0.7	0.8	1.2	0.5
Förslitningsskillnad, $w_i = x_i - y_i$	0.1	0.2	-0.1	0.3	0

De uppmätta parvisa förslitningsskillnaderna w_1, \dots, w_n modelleras som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler W_1, \dots, W_n där μ mäter hur mycket mer i genomsnitt A-däck slits än B-däck.

Parametrarna μ och σ^2 skattas med

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i = 0.10 \quad \text{respektive} \quad s_w^2 = \frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2 = 0.025$$

så σ skattas med $s_w = \sqrt{0.025} = 0.1581$.

Med $n = 5$ observationer w_1, \dots, w_n och konfidensgrad $1 - \alpha = 0.95$ fås ur $t(n-1) = t(4)$ -tabell att $t_{\alpha/2} = t_{0.025} = 2.7764$, så det observerade intervallet för den genomsnittliga skillnaden i däckförslitning blir

$$\mu \in \bar{w} \pm t_{\alpha/2} \frac{s_w}{\sqrt{n}} = 0.10 \pm 0.1963 = [-0.0963, 0.2963] \quad (95\%).$$

12.25 a) Observationerna x_1, \dots, x_{n_1} av levervärden för personer utan medicinering modelleras som utfall av $N(\mu_x, \sigma)$ -fördelade stokastiska variabler. Observationerna y_1, \dots, y_{n_2} av levervärden för personer med behandling modelleras som utfall av $N(\mu_y, \sigma)$ -fördelade stokastiska variabler. Samtliga stokastiska variabler förutsätts vara oberoende.

Med $n_1 = 50$ och $n_2 = 25$ erhöj man

$$\bar{x} = 148.2 \quad \bar{y} = 151.7 \quad s_x = 10.0 \quad s_y = 8.0$$

så skillnaden $\mu_y - \mu_x$ skattas med $\bar{y} - \bar{x} = 3.5$.

Standardavvikelsen σ skattas med s där

$$s^2 = \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{(n_1 - 1) + (n_2 - 1)} = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2} = 88.164$$

så $s = 9.3896$. Eftersom $D(\bar{Y} - \bar{X}) = \sigma \sqrt{\frac{1}{n_2} + \frac{1}{n_1}}$ så är medelfelet $d(\bar{Y} - \bar{X}) = s \sqrt{\frac{1}{n_2} + \frac{1}{n_1}} = 2.3$.

Ur $t(n_1 + n_2 - 2) = t(73)$ -tabeller fås $t_{0.025} = 1.993$ och ett 95% konfidensintervall för $\mu_y - \mu_x$ ges av

$$\mu_y - \mu_x \in (\bar{y} - \bar{x}) \pm t_{0.025} d(\bar{Y} - \bar{X}) = 3.5 \pm 1.993 \cdot 2.3 = 3.5 \pm 4.5 = [-1.08, 8.08] \quad (95\%).$$

b) Observationerna x_1, \dots, x_n av levervärden för personer före medicinering och y_1, \dots, y_n motsvarande efter medicinering, sammanfattas av

$$n = 25 \quad \bar{x} = 149.0 \quad \bar{y} = 150.9.$$

Förändringarna $z_i = y_i - x_i$ modelleras av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler Z_1, \dots, Z_n . Den genomsnittliga höjningen μ skattas med $\bar{z} = \bar{y} - \bar{x} = 1.9$ och σ skattas med $s_z = 1.6$.

Eftersom $D(\bar{Z}) = \sigma/\sqrt{n}$ så är medelfelet $d(\bar{Z}) = s_z/\sqrt{n} = 0.32$.

Ur $t(n-1) = t(24)$ -tabeller fås $t_{0.025} = 2.064$ och ett 95% konfidensintervall för μ ges av

$$\mu \in \bar{z} \pm t_{0.025} d(\bar{Z}) = 1.9 \pm 2.064 \cdot 0.32 = 1.9 \pm 0.66 = [1.24, 2.56] \quad (95\%).$$

12.26 Observationerna x_1, \dots, x_{n_1} av titanoxidhalten i det väl homogeniserade partiet modelleras som utfall av $N(\mu_1, \sigma_1)$ -fördelade stokastiska variabler där $\sigma_1 = 0.02$. Observationerna y_1, \dots, y_{n_2} av motsvarande halter i ett sämre homogeniserat parti modelleras som utfall av $N(\mu_2, \sigma_2)$ -fördelade stokastiska variabler där $\sigma_2 = 0.05$. Samtliga stokastiska variabler förutsätts vara oberoende.

Med $n_1 = 10$ och $n_2 = 50$ erhöj man $\bar{x} = 0.51$ och $\bar{y} = 0.69$ så skillnaden $\mu_2 - \mu_1$ skattas med $\bar{y} - \bar{x} = 0.18$. Skattningen är ett utfall av $\bar{Y} - \bar{X}$, där $\bar{Y} - \bar{X}$ är $N\left(\mu_2 - \mu_1, \sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}\right)$.

Ur $N(0, 1)$ -tabeller fås $\lambda_{0.05} = 1.6449$ och ett 90% konfidensintervall för $\mu_2 - \mu_1$ ges av

$$\mu_2 - \mu_1 \in (\bar{y} - \bar{x}) \pm \lambda_{0.05} D(\bar{Y} - \bar{X}) = 0.18 \pm 1.6449 \cdot 0.009487 = 0.18 \pm 0.0156 = [0.16, 0.20].$$

12.27 Låt x_1, \dots, x_{n_1} vara utfall av $N(\mu_1, \sigma)$ -fördelade stokastiska variabler och y_1, \dots, y_{n_2} vara utfall av $N(\mu_2, \sigma)$ -fördelade stokastiska variabler. Alla stokastiska variabler antas vara oberoende.

Skillnaden $\mu_1 - \mu_2$ skattas med $\bar{x} - \bar{y} = 49.2 - 37.4 = 11.8$. Vidare är $D(\bar{X} - \bar{Y}) = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ där σ skattas med

$$s = \sqrt{\frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{(n_1 - 1) + (n_2 - 1)}} = \sqrt{\frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2}} = 2.20.$$

Alltså är medelfelet $d(\bar{X} - \bar{Y}) = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.10$. Ur $t(n_1 + n_2 - 2) = t(16)$ -tabeller fås $t_{0.05} = 1.7459$ och ett 90% konfidensintervall för $\mu_1 - \mu_2$ ges av

$$\mu_1 - \mu_2 \in \bar{x} - \bar{y} \pm t_{0.05} d(\bar{X} - \bar{Y}) = 11.8 \pm 1.7459 \cdot 1.10 = 11.8 \pm 1.92 = [9.88, 13.7] \quad (90\%).$$

12.28 Observationerna x_1, \dots, x_{n_1} är utfall av $N(\mu_1, \sigma_1)$ -fördelade stokastiska variabler där σ_1 är känd.

Ett 95% konfidensintervall för μ_1 baserat på $n_1 = 5$ observationer ges av

$$\mu_1 \in \bar{x} \pm \lambda_{0.025} \frac{\sigma_1}{\sqrt{n_1}} = \bar{x} \pm 1.96 \frac{\sigma_1}{\sqrt{5}} = (1.37, 1.53) = 1.45 \pm 0.08$$

vilket ger $\bar{x} = 1.45$ och $\sigma_1 = 0.08\sqrt{5}/1.96 = 0.091268$.

Observationerna y_1, \dots, y_{n_2} är utfall av $N(\mu_2, \sigma_2)$ -fördelade stokastiska variabler där σ_2 är känd.

Ett 95% konfidensintervall för μ_2 baserat på $n_2 = 7$ observationer ges av

$$\mu_2 \in \bar{y} \pm \lambda_{0.025} \frac{\sigma_2}{\sqrt{n_2}} = \bar{y} \pm 1.96 \frac{\sigma_2}{\sqrt{7}} = (1.17, 1.29) = 1.23 \pm 0.06$$

vilket ger $\bar{y} = 1.23$ och $\sigma_2 = 0.06\sqrt{7}/1.96 = 0.080992$.

Skillnaden $\mu_1 - \mu_2$ skattas med $\bar{x} - \bar{y} = 0.22$. Skattningen är ett utfall av $\bar{X} - \bar{Y}$, där $\bar{X} - \bar{Y}$ är $N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$.

Ett 95% konfidensintervall för $\mu_1 - \mu_2$ ges av

$$\mu_1 - \mu_2 \in (\bar{x} - \bar{y}) \pm \lambda_{0.025} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0.22 \pm 1.96 \cdot 0.05102 = 0.22 \pm 0.10 = [0.12, 0.32].$$

12.29 Låt x_1, \dots, x_n vara studietiderna för de $n = 25$ personerna som undersökts. Dessa modelleras som utfall av oberoende och likafördelade stokastiska variabler X_1, \dots, X_n med $E(X) = \mu$ och $D(X) = \sigma$. Väntevärdet skattas med $\bar{x} = 49.3$ och σ med $s = 9.3$. Skattningen \bar{x} har medelfel $s/\sqrt{n} = 1.86$.

Med Centrala gränsvärdessatsen är $\sum_{i=1}^n X_i$ approximativt normalfördelat och således är \bar{X} approximativt $N(\mu, \sigma/\sqrt{n})$. Ett konfidensintervall för μ ges av

$$\mu \in \bar{x} \pm \lambda_{0.025} \frac{s}{\sqrt{n}} = 49.3 \pm 1.96 \cdot 1.86 = 49.3 \pm 3.65 = [45.654, 52.946] \quad (\approx 95\%).$$

12.30 Låt x_1, \dots, x_n , $n = 30$, vara de uppmätta föroreningshalterna uppströms och y_1, \dots, y_r , $r = 40$, motsvarande nedströms. Vi antar modellen att x_1, \dots, x_n är utfall av likafördelade stokastiska variabler X_1, \dots, X_n med $E(X) = \mu_x$ och $D(X) = \sigma_x$. På motsvarande sätt är y_1, \dots, y_r utfall av likafördelade Y_1, \dots, Y_r där $E(Y) = \mu_y$ och $D(Y) = \sigma_y$. Alla stokastiska variabler förutsätts vara oberoende.

Skillnaden $\mu_y - \mu_x$ skattas med $\bar{y} - \bar{x} = 86.1 - 13.2 = 72.9$. Skattningen $\bar{y} - \bar{x}$ är ett utfall av en stokastisk variabel $\bar{Y} - \bar{X}$ med standardavvikelse $D(\bar{Y} - \bar{X}) = \sqrt{\frac{\sigma_y^2}{r} + \frac{\sigma_x^2}{n}}$. Alltså har skattningen $\bar{y} - \bar{x}$ medelfel

$$d(\bar{Y} - \bar{X}) = \sqrt{\frac{s_y^2}{r} + \frac{s_x^2}{n}} = \sqrt{\frac{(38.7)^2}{40} + \frac{(2.80)^2}{30}} = 6.1403.$$

Med Centrala gränsvärdesatsen är $\sum_{i=1}^n X_i$ och $\sum_{i=1}^r Y_i$ approximativt normalfördelade och således är $\bar{Y} - \bar{X}$ approximativt $N\left(\mu_y - \mu_x, \sqrt{\frac{\sigma_y^2}{r} + \frac{\sigma_x^2}{n}}\right)$. Ett konfidensintervall för skillnaden $\mu_y - \mu_x$ ges av

$$\mu_y - \mu_x \in \bar{y} - \bar{x} \pm \lambda_{0.025} \sqrt{\frac{s_y^2}{r} + \frac{s_x^2}{n}} = 72.9 \pm 1.96 \cdot 6.1403 = 72.9 \pm 12.04 = [60.87, 84.94] \quad (\approx 95\%).$$

12.31 Av $n = 600$ undersökta enheter befanns $x = 24$ vara felaktiga. Om partiet är mycket stort kan vi modellera x som ett utfall av en binomialfördelad stokastisk variabel X , det vill säga X är $\text{Bin}(n, p)$ där p är andelen felaktiga enheter i partiet.

Vi skattar p med $p_{\text{obs}}^* = x/n = 0.04$. Skattningen p_{obs}^* beskrivs av $p^* = X/n$ som har

$$V(p^*) = V\left(\frac{1}{n}X\right) = \frac{1}{n^2} \underbrace{V(X)}_{np(1-p)} = \frac{p(1-p)}{n}$$

och standardavvikelse $D(p^*) = \sqrt{p(1-p)/n}$. Medelfelet för skattningen p_{obs}^* är

$$d(p^*) = \sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n}} = \sqrt{\frac{0.040(1-0.040)}{600}} = 0.008.$$

Eftersom $V(X) = np(1-p)$ skattas med $np_{\text{obs}}^*(1-p_{\text{obs}}^*) = 23.04$ vilket är större än 10 kan binomialfördelningen approximeras med normalfördelningen. Alltså är även $p^* = X/n$ approximativt normalfördelad. Med $1 - \alpha = 0.95$ är $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ och ett konfidensintervall för p ges av

$$p \in p_{\text{obs}}^* \pm \lambda_{0.025} d(p^*) = 0.040 \pm 1.96 \cdot 0.008 = 0.040 \pm 0.0157 = [0.024, 0.056] \quad (\approx 95\%).$$

12.32 Låt X vara $\text{Bin}(n, p)$ och $p^* = X/n$ vara den stokastiska variabel som beskriver skattningen $p_{\text{obs}}^* = x/n$ av p . Antag att $V(X)$ är så stor att normalapproximation av binomialfördelningen är tillämplig. Då ges ett approximativt konfidensintervall för p av

$$p \in p_{\text{obs}}^* \pm \lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} = p^* \pm \Delta(p)$$

där $\Delta(p) = \lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$ är halva bredden av konfidensintervallet. Bredden kommer således att bero på p . Lösning av

$$\frac{d}{dp} [p(1-p)] = 1 - 2p = 0$$

ger $p = 1/2$ och kontroll av andraderivatans ger att detta är ett maximum. Alltså är $p(1-p) \leq \frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$ för $0 \leq p \leq 1$ och

$$\Delta(p) = \lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq \lambda_{\alpha/2} \sqrt{\frac{1}{4n}} = \Delta.$$

Omformat fås att

$$n = \frac{\lambda_{\alpha/2}^2}{4\Delta^2}$$

observationer ger ett konfidensintervall vars halva bredd $\Delta(p)$ är högst Δ .

För att erhålla ett 95% konfidensintervall sådant att $|p - p_{\text{obs}}^*| \leq 0.005$, det vill säga $p \in p_{\text{obs}}^* \pm 0.005$ väljs $\Delta = 0.005$ och $1 - \alpha = 0.95$. Då är $\lambda_{\alpha/2} = 1.9600$ och

$$n = \frac{1.96^2}{4 \cdot 0.005^2} = 38416$$

dvs med minst 38416 observationer fås ett konfidensintervall med önskad precision.

Med kunskap om att $p \leq 0.04$ kan man utnyttja att $p(1-p)$ är växande på intervallet $[0, \frac{1}{2}]$ och

$$\Delta(p) \leq \lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \Bigg|_{p=0.04} = \lambda_{\alpha/2} \sqrt{\frac{0.0384}{n}} = \Delta.$$

Omformat fås att

$$n = \frac{\lambda_{\alpha/2}^2 \cdot 0.0384}{\Delta^2} = \frac{1.96^2 \cdot 0.0384}{0.005^2} = 5900.7$$

dvs 5901 observationer ger ett konfidensintervall vars halva bredd $\Delta(p)$ är högst $\Delta = 0.005$ då $p \leq 0.04$.

(Med 5901 observationer måste $p \geq 0.0016975$ för att $V(X) \geq 10$ och normalapproximationen vara tillåten.)

12.33 Låt x_1 och x_2 vara antalet sympatisörer med det borgerliga blocket i oktober och november. Vi modellerar x_1 och x_2 som utfall av oberoende binomialfördelade stokastiska variabler X_1 och X_2 , där X_1 är $\text{Bin}(n_1, p_1)$ och X_2 är $\text{Bin}(n_2, p_2)$. Förändringen $p_2 - p_1$ skattas med $(p_2)_{\text{obs}}^* - (p_1)_{\text{obs}}^* = 0.456 - 0.465 = -0.009$. Eftersom

$$V(p_2^* - p_1^*) = \{\text{ober.}\} = V(p_2^*) + (-1)^2 V(p_1^*) = \frac{1}{n_2} V(X_2) + \frac{1}{n_1} V(X_1) = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$$

fås medelfelet för $(p_2)_{\text{obs}}^* - (p_1)_{\text{obs}}^*$ till

$$\begin{aligned} d(p_2^* - p_1^*) &= \sqrt{\frac{(p_2)_{\text{obs}}^*(1-(p_2)_{\text{obs}}^*)}{n_2} + \frac{(p_1)_{\text{obs}}^*(1-(p_1)_{\text{obs}}^*)}{n_1}} = \sqrt{\frac{0.456(1-0.456)}{1689} + \frac{0.465(1-0.465)}{1704}} \\ &= 0.017113. \end{aligned}$$

Varianserna $V(X_1) = n_1 p_1(1-p_1)$ och $V(X_2) = n_2 p_2(1-p_2)$ skattas med $n_1 (p_1)_{\text{obs}}^*(1-(p_1)_{\text{obs}}^*) = 423.91$ respektive $n_2 (p_2)_{\text{obs}}^*(1-(p_2)_{\text{obs}}^*) = 418.98$. Eftersom båda är större än 10 (med råge!) kan binomialfördelningarna approximeras med normalfördelningar och även $p_2^* - p_1^*$ kan antas vara approximativt normalfördelad.

Med $1 - \alpha = 0.95$ fås $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ och ett konfidensintervall för förändringen $p_2 - p_1$ ges av

$$p_2 - p_1 \in (p_2)_{\text{obs}}^* - (p_1)_{\text{obs}}^* \pm \lambda_{0.025} d(p_2^* - p_1^*) = -0.009 \pm 1.96 \cdot 0.017 = -0.009 \pm 0.0335 \quad (\approx 95\%).$$

12.34 Låt $x = 36$ vara antalet defekta bland de $n = 1000$ undersökta enheterna. Vi modellerar x som ett utfall av en hypergeometrisk fördelad stokastisk variabel X , X är $\text{Hypergeo}(N, n, p)$ där $N = 100000$ och p är andelen defekta.

Andelen p skattas med $p_{\text{obs}}^* = x/n = 0.036$. Eftersom $V(X) = np(1-p)(N-n)/(N-1)$ skattas till

$$np_{\text{obs}}^*(1-p_{\text{obs}}^*) \frac{N-n}{N-1} = 1000 \cdot 0.036(1-0.036) \cdot 0.99 = 34.357$$

vilket är större än 10 kan den hypergeometriska fördelningen approximeras med en normalfördelning och även $p^* = X/n$ är approximativt normalfördelat. Medelfelet för p_{obs}^* är

$$d(p^*) = \sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n} \cdot \frac{N-n}{N-1}} = \sqrt{\frac{0.036(1-0.036)}{1000}} \cdot 0.99 = 0.0058615.$$

Med $1 - \alpha = 0.95$ fås $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ och ett konfidensintervall för andelen defekta p ges av

$$p \in p_{\text{obs}}^* \pm \lambda_{0.025} d(p^*) = 0.036 \pm 1.96 \cdot 0.0058615 = 0.036 \pm 0.0115 \quad (\approx 95\%).$$

b) Det totala antalet defekta i partiet Np skattas med $Np_{\text{obs}}^* = 3600$. Denna skattning har medelfel

$$d(Np^*) = N \sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n} \cdot \frac{N-n}{N-1}} = 586.15$$

Ett konfidensintervall för antalet defekta ges av

$$Np \in Np_{\text{obs}}^* \pm \lambda_{0.025} d(Np^*) = 3600 \pm 1.96 \cdot 586.15 = 3600 \pm 1149 \quad (\approx 95\%).$$

12.35 Låt X vara $\text{Hyp}(N, n, p)$ och $p^* = X/n$ vara den stokastiska variabel som beskriver skattningen $p_{\text{obs}}^* = x/n$ av p . Antag att $V(X)$ är så stor att normalapproximation av den hypergeometriska fördelningen är tillämplig. Då ges ett approximativt konfidensintervall för p av

$$p \in p_{\text{obs}}^* \pm \lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n} \frac{N-n}{N-1}}$$

Om forskare F' skall få ett konfidensintervall av samma bredd skall n' väljas så att

$$\lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n'} \frac{N'-n'}{N'-1}} = \lambda_{\alpha/2} \sqrt{\frac{p(1-p)}{n} \frac{N-n}{N-1}}$$

det vill säga

$$n' = \frac{N'}{\frac{1}{n} \frac{N-n}{N-1} (N'-1) + 1} = 224.89.$$

Alltså undersökningen bör omfatta 225 personer.

För forskaren F'' bör urvalsstorleken

$$n'' = \frac{N''}{\frac{1}{n} \frac{N-n}{N-1} (N''-1) + 1} = 249.13.$$

vara minst 250 personer.

12.36 Låt $x = 400$ vara antalet bilar som passerar under ett tidsintervall av längd $t = 10$. Vi ansätter modellen att x är ett utfall av X som är poissonfördelat med väntevärde $\lambda t = 10\lambda$. Om $\lambda t \geq 15$ så är X approximativt $N(\lambda t, \sqrt{\lambda t})$. Här skattas väntevärdet λt med $x = 400$ så vi antar att normalapproximation är tillämplig. Ett konfidensintervall för λt med konfidensgrad $\approx 1 - \alpha = 0.95$ ges av

$$10\lambda \in x \pm \lambda_{0.025} d(x) = 400 \pm 1.96\sqrt{400} = 400 \pm 39.2 = (361, 439).$$

Ett konfidensintervall för λ blir således

$$\lambda \in \frac{400}{10} \pm \frac{1.96\sqrt{400}}{10} = 40 \pm 3.92 = (36.1, 43.9). \quad (\approx 95\%)$$

12.37 Låt x_1, \dots, x_n vara antal samtal under det aktuella tidsintervallet för de olika dagarna. Dessa modelleras som utfall av oberoende Poisson(μ)-fördelade stokastiska variabler X_1, \dots, X_n . Parametern μ (väntevärdet) skattas med

$$\mu_{\text{obs}}^* = \bar{x} = 100.88 \text{ samtal.}$$

Skattningen beskrivs av

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

där

$$Y = X_1 + \dots + X_n$$

är Poissonfördelad med väntevärde

$$E(Y) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = n\mu$$

som skattas med $\sum x_i = 807$ vilket är större än 15 med råge. Alltså kan Poissonfördelningen för Y approximeras med normalfördelning, men då är även \bar{X} approximativt normalfördelad. Alltså,

$$\bar{X} \text{ är approximativt } N\left(\mu, \sqrt{\frac{\mu}{n}}\right)$$

Ett konfidensintervall för m med konfidensgrad approximativt $1 - \alpha = 0.95$ ges av

$$\mu \in \bar{x} \pm \lambda_{\alpha/2} \sqrt{\frac{\bar{x}}{n}} = 100.88 \pm 6.9599 = [93.9, 107.8] \quad (\approx 95\%).$$

12.38 Med $X_- = \min(X_1, \dots, X_n)$ som ett minimum av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n . Då är

$$\begin{aligned} P(X_- \geq \mu) &= P(\min(X_1, \dots, X_n) \geq \mu) = P(X_1 \geq \mu, \dots, X_n \geq \mu) = \{\text{oberoende}\} \\ &= P(X_1 \geq \mu) \cdots P(X_n \geq \mu) = (1 - \Phi(0))^n = \left(\frac{1}{2}\right)^n. \end{aligned}$$

På samma sätt är

$$P(X_+ \leq \mu) = P(\max(X_1, \dots, X_n) \leq \mu) = P(X_1 \leq \mu) \cdots P(X_n \leq \mu) = \left(\frac{1}{2}\right)^n$$

så

$$P(X_- \leq \mu \leq X_+) = 1 - P(\{X_- \geq \mu\} \cup \{X_+ \leq \mu\}) = 1 - \frac{1}{2^n} - \frac{1}{2^n} = 1 - \frac{1}{2^{n-1}}.$$

Alltså är (x_-, x_+) ett konfidensintervall för μ med konfidensgrad $1 - \frac{1}{2^{n-1}}$.

12.39 Från uppgift 8.7 hämtas de $n = 20000$ simulerade observationerna x_1, \dots, x_n som är utfall av Frechét-fördelade stokastiska variabler (med väntevärde 1.0655).

Då n är stort är enligt Centrala gränsvärdessatsen $\sum_{i=1}^n x_i$ ett utfall av en approximativt normalfördelad stokastisk variabel och så även $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.0653$.

Variansen $\sigma^2 = V(X)$ skattas med $s^2 = 0.0201$. Ett konfidensintervall för $\mu = E(X)$, med konfidensgrad $\approx 1 - \alpha = 95\%$, ges av

$$\mu \in \bar{x} \pm \lambda_{\alpha/2} \frac{s}{\sqrt{n}} = 1.0653 \pm 1.96 \frac{\sqrt{0.0201}}{\sqrt{20000}} = 1.065 \pm 0.002 = (1.063, 1.067) \quad (\approx 95\%).$$

Notera att detta konfidensintervall kom att täcka väntevärdet.

12.40 Från uppgift 8.6 har vi att

```
>> z = sum(2*x <= y.^2)
z =
    2340
```

gav en skattning av p som var $p_{\text{obs}}^* = 2340/10000 = 0.2340$. Antalet par, z , sådana att $2x \leq y^2$ är ett utfall av en stokastisk variabel Z där Z är $\text{Bin}(n, p)$ med $n = 10000$. Variansen $V(Z) = np(1-p)$ skattas med $np_{\text{obs}}^*(1-p_{\text{obs}}^*) = 1792.4$ vilket är större än 10 med råge så binomialfördelningen kan approximeras med en normalfördelning vilket ger att även $p^* = Z/n$ är approximativt normalfördelad, p^* är approximativt $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

Ett konfidensintervall för p , med konfidensgrad $\approx 1 - \alpha = 99\%$, ges av

$$p \in p_{\text{obs}}^* \pm \lambda_{\alpha/2} \sqrt{\frac{p_{\text{obs}}^*(1-p_{\text{obs}}^*)}{n}} = 0.2340 \pm 2.5758 \cdot 0.0042337 = 0.234 \pm 0.011 = (0.223, 0.245).$$

Notera att detta konfidensintervall kom att täcka $p = \frac{1}{3\sqrt{2}} = 0.2357$.

13.1 Låt x vara antalet gånger som Pål svarar rätt.

a) Om Pål chansar ansätter vi modellen att är x ett utfall av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(n, p)$ med $n = 15$ och $p = 0.5$.

b) I binomialfördelningen är

$$P(X \geq 11) = \sum_{k=11}^{15} \binom{n}{k} p^k (1-p)^{n-k} = 0.05923.$$

Om man har tillgång till fördelningsfunktionen för binomialfördelningen kan sannolikheten erhållas enligt

$$P(X \geq 11) = 1 - P(X < 11) = 1 - P(X \leq 10) = 1 - F_X(10) = 1 - 0.94077 = 0.05923.$$

13.2 Om Pål chansar ansätter vi modellen att antalet rätta svar, x , är ett utfall av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(n, p)$ med $p = 0.5$.

Per förkastar hypotesen att Pål chansar om utfallet n observeras. Då Pål chansar är sannolikheten för detta

$$\alpha = P(X = n) = \binom{n}{n} p^n (1-p)^{n-n} = \frac{1}{2^n}.$$

För att få signifikansnivå $\alpha \leq 10^{-6}$ skall n väljas så att $\frac{1}{2^n} \leq 10^{-6}$ eller $n \geq 6 \ln(10) / \ln(2) = 19.9$.

Alltså, gör $n = 20$ försök. Signifikansnivån blir då $1/2^n \approx 9.54 \cdot 10^{-7}$.

13.3 Om Petter har ESP ansätter vi modellen att antalet rätta svar, x , är ett utfall av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(n, p)$ med $p = 0.80$. Om Petter har ESP är chansen att Per blir övertygad

$$P(X = n) = \binom{n}{n} p^n (1-p)^{n-n} = 0.8^{20} = 0.0115.$$

13.4 Låt X beskriva en glödlampas lystid i timmar. Med modellen att X är exponentialfördelad med väntevärde $\theta > 0$ har X täthetsfunktion

$$f_X(x) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0.$$

Fördelningsfunktionen för X blir då för $t > 0$

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx = \int_0^t \frac{1}{\theta} e^{-x/\theta} dx = 1 - e^{-t/\theta}.$$

a) Ekvationen $\alpha = P(X < a) = F_X(a) = 1 - e^{-a/\theta}$ ger $a = -\theta \ln(1 - \alpha)$ och med $\theta = 1000$ erhålls $a = -1000 \ln(1 - \alpha)$.

b) Med $\alpha = 0.05$ är $a = -1000 \ln(1 - 0.05) = 51.293$. Testet som förkastar hypotesen $\theta = 1000$ till förmån för hypotesen $\theta < 1000$ om $x < 51.293$ har signifikansnivå 5%. Här är $x = 75$ och hypotesen $\theta = 1000$ förkastas ej.

c) Med $x = 50 < 51.293$ förkastas hypotesen $\theta = 1000$ på nivå 5% till förmån för hypotesen $\theta < 1000$.

d) Om a väljes så att $a = x = 45$ är

$$P(X < a) = F_X(a) = 1 - e^{-a/\theta} = 1 - e^{-45/1000} = 0.044003.$$

Alltså, den lägsta signifikansnivå hypotesen $\theta = 1000$ kan förkastas på, givet observationen $x = 45$, är 4.4%.

13.5 Låt x vara antalet spel till första vinst. Vi modellerar x som ett utfall av den ffg(p)-fördelade stokastiska variabeln X , det vill säga

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

Då är

$$P(X > k) = (1 - p)^k, \quad k = 0, 1, 2, 3, \dots$$

Med $H_0 : p = 0.2$ och $H_1 : p < 0.2$ så förkastas H_0 till förmån för H_1 för stora värden på x . Med förkastelseområde $\{x, x + 1, x + 2, \dots, x\}$ är signifikansnivån för testet

$$P(X \geq x | H_0) = P(X > x - 1 | H_0) = (1 - p)^{x-1} \Big|_{p=0.20} = (0.80)^{x-1}.$$

Med utfallet $x = 11$ förkastas H_0 på (lägsta) signifikansnivå

$$P(X \geq 11 | H_0) = P(X > 10 | H_0) = (0.80)^{10} = 0.1074.$$

Alltså, vi förkastar inte H_0 på nivå 10%.

13.6 Med x_1, \dots, x_n som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n är \bar{x} ett utfall av \bar{X} som är $N(\mu, \sigma/\sqrt{n})$ -fördelad. Ett test av

$$H_0 : \mu = 0.5 \text{ (oskyldig)} \quad \text{mot} \quad H_1 : \mu > 0.5 \text{ (skyldig)}$$

görs på signifikansnivå $\alpha = 1\%$ om H_0 förkastas då $\bar{x} > 0.5 + \lambda_{0.01} \sigma / \sqrt{n}$. Eftersom signifikansnivån begränsar sannolikheten för fel av första slaget, $P(\text{Förkasta sann } H_0) \leq \alpha$, kommer sannolikheten för att en oskyldig person förklaras skyldig vara högst 1%, dvs alternativ 2 är en riktig tolkning.

I alternativ 3 beskrivs sannolikheten för fel av andra slaget, det vill säga P (ej förkasta falsk H_0) och den begränsas inte av signifikansnivån. Alternativ 1 och 4 ger betingade fördelningar på rummet {skyldig, oskyldig}, vilket inte modelleras i uppgiften.

13.7 Vi modellerar de uppmätta planklängderna x_1, \dots, x_n som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n . Vi vill testa hypotesen

$$H_0 : \sigma = 0.4 \quad \text{mot} \quad H_1 : \sigma \neq 0.4$$

på signifikansnivå $\alpha = 0.05$.

Alternativ 1. Med skattningen $s^2 = 0.46962$ fås ett konfidensintervall för σ med konfidensgrad $1 - \alpha = 0.95$ av (uppgift 12.18b)

$$0.50623 = \sqrt{\frac{(n-1)s^2}{27.5}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{6.26}} = 1.0606 \quad (95\%).$$

Eftersom 0.4 inte ingår i intervallet förkastas $H_0 : \sigma = 0.4$ till förmån för H_1 på risknivå $\alpha = 0.05$.

Alternativ 2. Vi förkastar H_0 då s är mycket större eller mycket mindre än 0.4, alternativt då

$$\frac{(n-1)s^2}{0.4^2}$$

är väldigt stor eller väldigt liten. Då H_0 är sann så är

$$\frac{(n-1)S^2}{0.4^2} \text{ en } \chi^2(n-1)\text{-fördelad stokastisk variabel.}$$

Alltså, med $\alpha = 0.05$ så är

$$P\left(6.26 \leq \frac{(n-1)S^2}{0.4^2} \leq 27.5\right) = 0.95.$$

Vi förkastar H_0 till förmån för H_1 om vi observerar att $(n-1)s^2/0.4^2$ mindre än 6.26 eller större än 27.5. Här:

$$\frac{(n-1)s^2}{0.4^2} = 44.027$$

så H_0 förkastas på nivå $\alpha = 0.05$.

13.8 Vi modellerar den uppmätta grumligheten x_1, \dots, x_n som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n där $\sigma = 0.2$. Vi vill testa hypotesen

$$H_0 : \mu = 4.0 \quad \text{mot} \quad H_1 : \mu > 4.0$$

på signifikansnivå $\alpha = 0.05$.

Alternativ 1: Vi förkastar H_0 för stora värden på \bar{x} , alternativt, då $(\bar{x} - 4)/(\sigma/\sqrt{n})$ är stor. Då H_0 är sann är

$$\frac{\bar{X} - 4}{\sigma/\sqrt{n}} \sim N(0, 1)$$

så $P((\bar{X} - 4)/(\sigma/\sqrt{n}) > \lambda_{0.05} | H_0) = 0.05$.

Alltså, H_0 förkastas om $(\bar{x} - 4)/(\sigma/\sqrt{n}) > \lambda_{0.05}$. Här är $\lambda_{0.05} = 1.6449$ och $\bar{x} = 4.1$ och

$$\frac{\bar{x} - 4}{\sigma/\sqrt{n}} = \frac{4.1 - 4}{0.2/\sqrt{10}} = 1.5811 < \lambda_{0.05},$$

så H_0 förkastas ej på nivå 5%.

Alternativ 2: Vi förkastar H_0 för stora värden på \bar{x} . Då H_0 är sann är

$$P(\bar{X} > x | H_0) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{x - \mu}{\sigma/\sqrt{n}} \mid H_0\right) = 1 - \Phi\left(\frac{x - 4.0}{0.2/\sqrt{10}}\right).$$

Med observationen $\bar{x} = 4.1$ fås att den lägsta signifikansnivån H_0 kan förkastas på är

$$P(\bar{X} > 4.1 | H_0) = 1 - \Phi\left(\frac{4.1 - 4.0}{0.2/\sqrt{10}}\right) = 1 - \Phi(1.5811) = 0.056923 > 0.05.$$

Alltså kan vi inte förkasta H_0 på nivå 5%.

13.9 Teststorheten

$$\frac{\bar{X} - 4}{\sigma/\sqrt{n}}$$

är normalfördelad med väntevärde

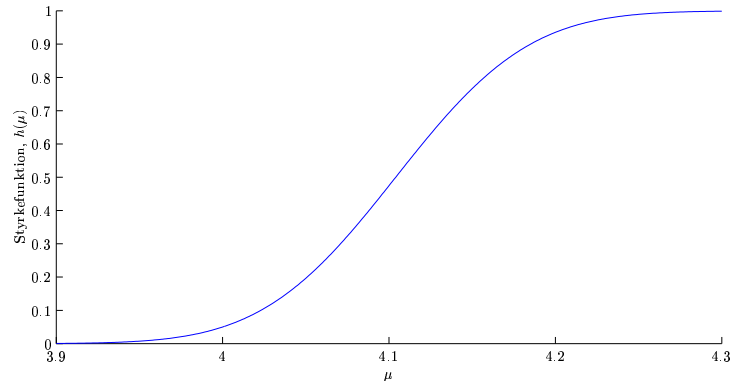
$$E\left(\frac{\bar{X} - 4}{\sigma/\sqrt{n}}\right) = \frac{E(\bar{X}) - 4}{\sigma/\sqrt{n}} = \frac{\mu - 4}{\sigma/\sqrt{n}}$$

och varians

$$V\left(\frac{\bar{X} - 4}{\sigma/\sqrt{n}}\right) = \frac{V(\bar{X})}{\sigma^2/n} = \frac{\sigma^2/n}{\sigma^2/n} = 1.$$

Styrkefunktionen blir då

$$\begin{aligned} h(\mu) &= P(\text{Förkasta } H_0) = P\left(\frac{\bar{X} - 4}{\sigma/\sqrt{n}} > \lambda_{0.05}\right) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \lambda_{0.05} + \frac{4 - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\lambda_{0.05} + \frac{4 - \mu}{\sigma/\sqrt{n}}\right). \end{aligned}$$



Styrkefunktionen $h(\mu)$, det vill säga sannolikheten att förkasta H_0 som funktion av μ .

$$\begin{aligned} h(3.8) &= 1 - \Phi\left(1.6449 + \frac{4-3.8}{0.2/\sqrt{10}}\right) = 1 - \Phi(4.8071) = 7.7 \cdot 10^{-7} \\ h(4.3) &= 1 - \Phi\left(1.6449 + \frac{4-4.3}{0.2/\sqrt{10}}\right) = 1 - \Phi(-3.0986) = 0.99903. \end{aligned}$$

13.10 Låt x_1, \dots, x_n vara de uppmätta smältpunkterna. Vi modellerar x_1, \dots, x_n som utfall av oberoende och likafördelade stokastiska variabler X_1, \dots, X_n , där

$$X = \text{smältpunkt} + \text{mätfel} = \mu + \epsilon$$

där ϵ är $N(0, \sigma)$, $\sigma = 2.3$. Då är $X_i \sim N(\mu, \sigma)$. Vi vill testa:

$$H_0 : \mu = \mu_0 = 1050^\circ \quad \text{mot} \quad H_1 : \mu \neq \mu_0$$

på nivå $\alpha = 0.05$.

Vi förkastar H_0 för stora värden på $|\bar{x} - \mu_0|$, alternativt stora värden på $|z|$ där

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

som då H_0 är sann är ett utfall på Z , en $N(0, 1)$ -fördelad stokastisk variabel.

Alternativ 1: Alltså, om H_0 är sann så är $P(|Z| > \lambda_{\alpha/2}) = \alpha$. Med $\alpha = 0.05$ är $\lambda_{\alpha/2} = 1.96$ och vi förkastar H_0 till förmån för H_1 om vi observerar ett utfall $|z| > 1.96$. Här fås

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1050.9 - 1050}{2.3/\sqrt{10}} = 1.2649$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

Alternativ 2: Vi observerar utfallet

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1050.9 - 1050}{2.3/\sqrt{10}} = 1.2649.$$

Om H_0 är sann är

$$P(|Z| > 1.2649) = 2(1 - \Phi(1.2649)) = 0.2059$$

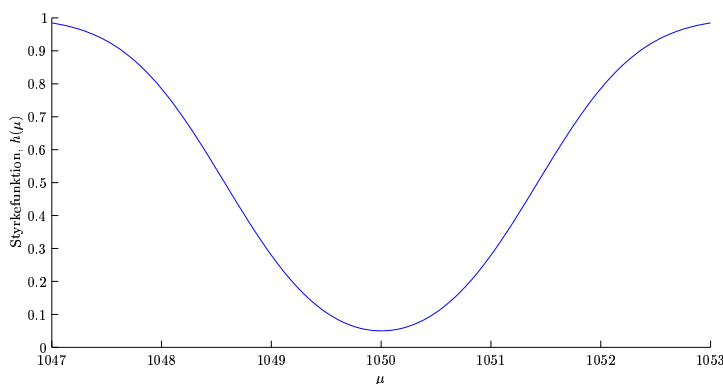
så den lägsta signifikansnivå vi skulle förkasta H_0 på är 0.2059, det vill säga vi förkastar inte H_0 på nivå 5%.

Styrkefunktionen $h(\mu)$ kan bestämmas som

$$\begin{aligned} h(\mu) &= P(\text{förkasta } H_0) = P\left(\left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| > \lambda_{\alpha/2}\right) = 1 - P\left(-\lambda_{\alpha/2} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}\right) \\ &= 1 - P\left(-\lambda_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \left(\Phi\left(\lambda_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(-\lambda_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)\right) \end{aligned}$$

Vi beräknar

$$h(1051) = 0.27968 \quad h(1053) = 0.9848.$$



Styrkefunktionen $h(\mu)$, det vill säga sannolikheten att förkasta $H_0 : \mu = 1050^\circ$ som funktion av μ .

13.11 Testet

$$H_0 : \mu = 1 \quad H_1 : \mu < 1$$

som förkastar H_0 då $\bar{x} < 1 - 2\lambda_\alpha/\sqrt{n}$ har signifikansnivå α då x_1, \dots, x_n är utfall av oberoende $N(\mu, 2)$ -fördelade stokastiska variabler.

Vi söker n så att

$$P(\bar{X} < 1 - 2\lambda_\alpha/\sqrt{n}) \geq 0.99$$

då $\mu = 0$. Då $\mu = 0$ är \bar{X} $N(0, 2/\sqrt{n})$ -fördelad och

$$0.99 \leq P\left(\bar{X} < 1 - \lambda_\alpha \frac{2}{\sqrt{n}}\right) = P\left(\frac{\bar{X} - 0}{2/\sqrt{n}} < \frac{1 - 0}{2/\sqrt{n}} - \lambda_\alpha\right) = \Phi\left(\frac{1}{2/\sqrt{n}} - \lambda_\alpha\right)$$

Men $0.99 = \Phi(\lambda_{0.01})$ så

$$\lambda_{0.01} \leq \frac{1}{2/\sqrt{n}} - \lambda_\alpha$$

eller

$$n \geq 4(\lambda_{0.01} + \lambda_\alpha)^2 = 4(2.3263 + 1.6449)^2 = 63.082.$$

Alltså, tag 64 eller fler observationer för att erhålla önskad styrka.

13.12 Om σ inte är känd så skattas den med stickprovsstandardavvikelsen $s = 2.028$.

Vi förkastar H_0 för stora värden på $|\bar{x} - \mu_0|$, alternativt stora värden på $|t|$ där

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$ -fördelad stokastisk variabel T . Alltså, om H_0 är sann så är $P(|T| > t_{\alpha/2}) = \alpha$. Med $\alpha = 0.05$ är $t_{\alpha/2} = 2.2622$ och vi förkastar H_0 till förmån för H_1 om vi observerar $|t| > 2.2622$. Här fås

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1050.9 - 1050}{2.028/\sqrt{10}} = 1.4345$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

13.13 Ur uppgift 12.24 fås att ett konfidensintervall för den genomsnittliga förslitningsskillnaden μ är

$$\mu \in \bar{w} \pm t_{\alpha/2} \frac{s_w}{\sqrt{n}} = 0.10 \pm 0.1963 = [-0.0963, 0.2963] \quad (95\%).$$

Eftersom $\mu = 0$ inte är ett orimligt värde på förslitningsskillnaden, 0 ingår i konfidensintervallet, kan inte hypotesen $H_0 : \mu = 0$ förkastas till förmån för $H_1 : \mu \neq 0$ på nivå 5%.

13.14 Mätningarna av tyngdkraftsaccelerationen x_1, \dots, x_n beskrivs av oberoende $N(g, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n . Skattningen av g , $\bar{x} = 972$ är ett utfall av \bar{X} som är en $N(g, \sigma/\sqrt{n})$ -fördelad stokastisk variabel.

Alternativ 1: Vi förkastar $H_0 : g = 981$ till förmån för $H_1 : g \neq 981$ för stora värden på $|t|$ där

$$t = \frac{\bar{x} - 981}{s/\sqrt{n}}$$

som om H_0 är sann är ett utfall av en $t(n-1) = t(17)$ -fördelad stokastisk variabel. Ur $t(17)$ -tabeller erhålls att $t_{\alpha/2} = t_{0.025} = 2.11$. Testet som förkastar H_0 till förmån för H_1 då $|t| > 2.11$ har signifikansnivå $\alpha = 0.05$. Här observeras utfallet

$$t = \frac{\bar{x} - 981}{s/\sqrt{n}} = \frac{972 - 981}{6.0/\sqrt{18}} = -6.364$$

så $|t| > 2.11$ och H_0 förkastas på nivå 5%.

Alternativ 2: Ett konfidensintervall för g ges av

$$g \in \bar{x} \pm t_{0.025} \frac{s}{\sqrt{n}} = 972 \pm 2.11 \frac{6.0}{\sqrt{18}} = 972 \pm 3 \quad (95\%).$$

Detta intervall innehåller inte $g = 981$ så H_0 förkastas till förmån för H_1 på risknivå 5%.

13.15 Låt x_1, \dots, x_n vara de uppmätta Hg-halterna i gäddorna. Dessa modelleras som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n .

a) Vi vill med $\mu_0 = 0.9$ testa hypotesen

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu > \mu_0$$

på nivå $\alpha = 0.05$.

Vi förkastar H_0 för stora värden på

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$ -fördelad stokastisk variabel T . Alltså, om H_0 är sann så är $P(T > t_\alpha) = \alpha$. Med $\alpha = 0.05$ är $t_\alpha = 1.8331$ och vi förkastar H_0 till förmån för H_1 om vi observerar $t > 1.8331$. Här fås

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.97 - 0.9}{0.33015/\sqrt{10}} = 0.67048$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

b) Vi vill med $\mu_0 = 1.1$ testa hypotesen

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu < \mu_0$$

på nivå $\alpha = 0.05$.

Vi förkastar H_0 för små värden på

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$ -fördelad stokastisk variabel T . Alltså, om H_0 är sann så är $P(T < t_{1-\alpha}) = \alpha$. Med $\alpha = 0.05$ är $t_{1-\alpha} = -1.8331$ och vi förkastar H_0 till förmån för H_1 om vi observerar $t < -1.8331$. Här fås

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.97 - 1.1}{0.33015/\sqrt{10}} = -1.2452$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

13.16 Observationerna x_1, \dots, x_n och y_1, \dots, y_n sammanfattas av tabellen:

Person, i	1	2	3	4	5	6	7	8
Morgon, x_i	172	168	180	181	160	163	165	177
Kväll, y_i	172	167	177	179	159	161	166	175
Skillnad, $w_i = x_i - y_i$	0	1	3	2	1	2	-1	2

De uppmätta parvisa längdskillnaderna w_1, \dots, w_n modelleras som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler W_1, \dots, W_n där μ mäter hur mycket längre i genomsnitt personerna är på morgonen än på kvällen.

Parametrarna μ och σ^2 skattas med

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i = 1.25 \quad \text{respektive} \quad s_w^2 = \frac{1}{n-1} \sum_{i=1}^n (w_i - \bar{w})^2 = 1.6429$$

så σ skattas med $s_w = \sqrt{1.6429} = 1.2817$.

Vi vill testa

$$H_0 : \mu = 0 \quad \text{mot} \quad H_1 : \mu \neq 0$$

på nivå $\alpha = 0.05$.

Alternativ 1: Vi förkastar H_0 för stora värden på $|t|$ där

$$t = \frac{\bar{w} - 0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$ -fördelad stokastisk variabel T . Alltså, om H_0 är sann så är $P(|T| > t_{\alpha/2}) = \alpha$. Med $\alpha = 0.05$ är $t_{\alpha/2} = 2.3646$ och vi förkastar H_0 till förmån för H_1 om vi observerar $|t| > 2.3646$. Här fås

$$t = \frac{\bar{w} - 0}{s/\sqrt{n}} = \frac{1.25 - 0}{1.2817/\sqrt{8}} = 2.7584$$

så H_0 förkastas till förmån för H_1 på nivå $\alpha = 0.05$.

Alternativ 2: Med $n = 8$ observationer w_1, \dots, w_n och konfidsensgrad $1 - \alpha = 0.95$ fås ur $t(n-1) = t(7)$ -tabell att $t_{\alpha/2} = t_{0.025} = 2.3646$, så konfidensintervallet för den genomsnittliga längdskillnaden blir

$$\mu \in \bar{w} \pm t_{\alpha/2} \frac{s_w}{\sqrt{n}} = 1.25 \pm 1.072 = [0.17844, 2.3216] \text{ cm} \quad (95\%).$$

Eftersom $\mu = 0$ inte är ett rimligt värde enligt konfidensintervallet förkastas hypotesen $H_0 : \mu = 0$ till förmån för $H_1 : \mu \neq 0$ på nivå 5%.

13.17 Låt x_1, \dots, x_n vara mätningarna av molekylvikt A och y_1, \dots, y_m motsvarande för B. Dessa värden modelleras som utfall av $N(\mu_1, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n respektive $N(\mu_2, \sigma)$ -fördelade stokastiska variabler Y_1, \dots, Y_m . Alla stokastiska variabler antas vara oberoende.

Vi vill testa:

$$H_0 : \mu_1 = \mu_2 \quad \text{mot} \quad H_1 : \mu_1 \neq \mu_2.$$

Dessa hypoteser formuleras som

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{mot} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Vi förkastar H_0 till förmån för H_1 för stora värden på $|t|$ där

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

och s är skattningen av σ som ges av

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{(n-1) + (m-1)} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2}.$$

Om H_0 är sann så är t ett utfall på en $t(n+m-2)$ -fördelad stokastisk variabel T . Ur $t(6+8-2) = t(12)$ -tabell fås att $P(|T| > t_{0.025}) = 0.05$ för $t_{0.025} = 2.1788$. Alltså, om H_0 förkastas då $|t| > 2.1788$ så har testet signifikansnivå 5%.

b) Med observationer

$$\begin{array}{l|cccccc} x_1, \dots, x_6 & 174.18 & 174.30 & 174.23 & 174.29 & 174.36 & 174.25 \\ y_1, \dots, y_8 & 174.19 & 174.40 & 174.20 & 174.35 & 174.32 & 174.14 & 174.27 & 174.34 \end{array}$$

fås

$$\bar{x} = 174.27, s_x = 0.062423 \quad \bar{y} = 174.28, s_y = 0.091486$$

vilket ger

$$s = 0.080659 \quad \text{och} \quad t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{174.27 - 174.28}{0.080659 \sqrt{\frac{1}{6} + \frac{1}{8}}} = -0.18174.$$

Eftersom $|t| < 2.1788$ kan inte H_0 förkastas på nivå 5%.

13.18 Mätningarna av tablettvikterna x_1, \dots, x_n beskrivs av oberoende och likafördelade stokastiska variabler X_1, \dots, X_n med väntevärde $\mu = E(X)$ och varians $\sigma^2 = V(X)$. Då n är stor är enligt Centrala gränsvärdesatsen $\sum_{i=1}^n x_i$ ett utfall av en approximativt normalfördelad stokastisk variabel och så även $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Alltså, skattningen av μ , $\bar{x} = 0.69$ är ett utfall av \bar{X} som är approximativt $N(\mu, \sigma/\sqrt{n})$ -fördelad.

Alternativ 1: Vi förkastar $H_0 : \mu = 0.65$ till förmån för $H_1 : \mu \neq 0.65$ för stora värden på $|z|$ där

$$z = \frac{\bar{x} - 0.65}{\sigma/\sqrt{n}}$$

som om H_0 är sann är ett utfall av en approximativt $N(0, 1)$ -fördelad stokastisk variabel. Ur $N(0, 1)$ -tabeller erhålls att $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Testet som förkastar H_0 till förmån för H_1 då $|z| > 1.96$ har signifikansnivå $\alpha \approx 0.05$. Här observeras utfallet

$$z = \frac{\bar{x} - 0.65}{\sigma/\sqrt{n}} = \frac{0.69 - 0.65}{0.02/\sqrt{35}} = 11.832$$

så $|z| > 1.96$ och H_0 förkastas på nivå $\approx 5\%$.

P-värdet för testet

$$P(|Z| > 11.832) \approx 2(1 - \Phi(11.832)) \approx 0$$

så hypotesen förkastas på alla rimliga signifikansnivåer.

Alternativ 2: Ett konfidensintervall för μ ges av

$$\mu \in \bar{x} \pm \lambda_{0.025} \frac{\sigma}{\sqrt{n}} = 0.69 \pm 1.96 \frac{0.02}{\sqrt{35}} = 0.69 \pm 0.007 = (0.683, 0.697) \quad (\approx 95\%).$$

Detta intervall innehåller inte $g = 0.65$ så H_0 förkastas till förmån för H_1 på risknivå $\approx 5\%$.

13.19 Mätningarna av tablettvikterna x_1, \dots, x_n beskrivs av oberoende och likafördelade stokastiska variabler X_1, \dots, X_n med väntevärde $\mu = E(X)$ och varians $\sigma^2 = V(X)$. Då n är stor är enligt Centrala gränsvärdesatsen $\sum_{i=1}^n x_i$ ett utfall av en approximativt normalfördelad stokastisk variabel och så även $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Alltså, skattningen av μ , $\bar{x} = 0.69$ är ett utfall av \bar{X} som är approximativt $N(\mu, \sigma/\sqrt{n})$ -fördelad där σ skattas med $s = 0.018$. Medelfelet $d(\bar{X})$ är en skattning av $D(\bar{X}) = \sigma/\sqrt{n}$, det vill säga $d(\bar{X}) = s/\sqrt{n} = 0.018/\sqrt{35} = 0.0030426$.

Alternativ 1: Vi förkastar $H_0 : \mu = 0.65$ till förmån för $H_1 : \mu \neq 0.65$ för stora värden på $|z|$ där

$$z = \frac{\bar{x} - 0.65}{s/\sqrt{n}}$$

som om H_0 är sann är ett utfall av en approximativt $N(0, 1)$ -fördelad stokastisk variabel. Ur $N(0, 1)$ -tabeller erhålls att $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Testet som förkastar H_0 till förmån för H_1 då $|z| > 1.96$ har signifikansnivå $\alpha \approx 0.05$. Här observeras utfallet

$$z = \frac{\bar{x} - 0.65}{s/\sqrt{n}} = \frac{0.69 - 0.65}{s/\sqrt{35}} = 13.147$$

så $|z| > 1.96$ och H_0 förkastas på nivå $\approx 5\%$.

P-värdet för testet

$$P(|Z| > 13.147) \approx 2(1 - \Phi(13.147)) \approx 0$$

så hypotesen förkastas på alla rimliga signifikansnivåer.

Alternativ 2: Ett konfidensintervall för μ ges av

$$\mu \in \bar{x} \pm \lambda_{0.025} \frac{s}{\sqrt{n}} = 0.69 \pm 1.96 \frac{0.018}{\sqrt{35}} = 0.69 \pm 0.006 = (0.684, 0.696) \quad (\approx 95\%).$$

Detta intervall innehåller inte $g = 0.65$ så H_0 förkastas till förmån för H_1 på risknivå 5% .

13.20 Låt x_1, \dots, x_n vara de uppmätta alkoholhalterna i flaskorna. Vi modellerar x_1, \dots, x_n som utfall av oberoende $N(\mu, \sigma)$ -fördelade stokastiska variabler X_1, \dots, X_n , där $\sigma = 0.10$. Vi vill testa:

$$H_0 : \mu = 3.0\% = \mu_0 \quad \text{mot} \quad H_1 : \mu < \mu_0$$

på nivå $\alpha = 0.01$. Signifikansnivån begränsar sannolikheten för att förkasta en korrekt nollhypotes, det vill säga att dra slutsatsen att en flaska har liten alkoholhalt fast den egentligen är hög.

Vi förkastar H_0 då för små värden på

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

som om H_0 är sann är ett utfall på en $N(0, 1)$ -fördelad stokastisk variabel Z .

Alternativ 1: Ur $N(0, 1)$ -tabell fås att $\lambda_{0.01} = 2.3263$ och $P(Z < -2.3263) = 0.01$. Vi observerar utfallet

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.98 - 3.0}{0.10/\sqrt{10}} = -0.632 > -\lambda_{0.01}.$$

Vi förkastar inte H_0 på nivå 1%.

Alternativ 2: Vi observerar utfallet

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.98 - 3.0}{0.10/\sqrt{10}} = -0.632.$$

Vi har att då H_0 är sann så är

$$P(\text{Förkasta } H_0) = P(Z < -0.632) = \Phi(-0.632) = 1 - \Phi(0.632) = 0.26354$$

vilket är den lägsta signifikansnivå H_0 kan förkastas på. Vi kan inte förkasta H_0 på nivå 1%.

13.21 Låt x_1, \dots, x_n och y_1, \dots, y_n vara de slumpmässiga stickproven av de 10 + 10 oberoende $N(\mu_1, \sigma_1)$ -respektive $N(\mu_2, \sigma_2)$ -fördelade stokastiska variablerna X_1, \dots, X_n och Y_1, \dots, Y_n , där $\sigma_1 = 0.3$ och $\sigma_2 = 0.4$.

Vi vill pröva hypotesen

$$H_0 : \mu_1 = \mu_2 \quad \text{mot} \quad H_1 : \mu_1 \neq \mu_2$$

på nivå $\alpha = 0.01$.

a) Vi förkastar H_0 för stora värden på $|z|$ där

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}$$

som om H_0 är sann är ett utfall av en $N(0, 1)$ -fördelad stokastisk variabel Z . På nivå $\alpha = 0.01$ förkastas H_0 om $|z| > \lambda_{0.01/2} = 2.5758$.

Om $\mu_1 - \mu_2 = 0.6$ är sannolikheten för att H_0 förkastas

$$\begin{aligned} P(|Z| > \lambda_{0.005}) &= 1 - P(-\lambda_{0.005} \leq Z \leq \lambda_{0.005}) = 1 - P\left(-\lambda_{0.005} \leq \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005}\right) \\ &= 1 - P\left(-\lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}\right) \\ &= 1 - P\left(-\lambda_{0.005} - 3.7947 \leq \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005} - 3.7947\right) \\ &= 1 - (\Phi(-1.2189) - \Phi(-6.3706)) = 0.8886. \end{aligned}$$

b) Ekvationen

$$1 - P \left(-\lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right) = 0.99$$

ger approximativt att

$$P \left(\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right) \approx 0.01$$

det vill säga

$$\lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \approx -\lambda_{0.01}$$

eller

$$\sqrt{n} \approx \frac{(\lambda_{0.01} + \lambda_{0.005})\sqrt{\sigma_1^2 + \sigma_2^2}}{\mu_1 - \mu_2} = \frac{(2.3263 + 2.5758)\sqrt{0.3^2 + 0.4^2}}{0.6} = 4.0851$$

eller $n \approx 16.68$. Alltså, minst 17 observationer krävs för att erhålla önskvärd styrka. (Styrkan är då 0.99115).

13.22 Om tärningen är rättvis skall antalet 6:or, x , av n tärningskast beskrivas av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(n, p)$ där $p = 1/6$.

Om tärningen misstänks ge för få sexor vill man testa hypotesen

$$H_0 : p = \frac{1}{6} \quad \text{mot} \quad H_1 : p < \frac{1}{6}$$

genom att förkasta för små värden på x .

a) Med utfallet $x = 0$ av $n = 12$ har testet p-värde

$$P(X \leq 0) = P(X = 0) = \binom{n}{0} p^0 (1-p)^{n-0} = \left(1 - \frac{1}{6}\right)^{12} = 0.11216$$

så H_0 förkastas ej på nivå 10%.

b) med utfallet $x = 8$ av $n = 120$ har testet p-värde

$$P(X \leq 8) = \sum_{k=0}^8 \binom{n}{k} p^k (1-p)^{n-k} = 0.0010151$$

så H_0 förkastas på nivå 1% men ej på nivå 0.1%.

Under nollhypotesen $H_0 : p = 1/6$ är $V(X) = np(1-p) = 16.667$ så man kan approximera binomialfördelningen med normalfördelningen. Då kan ovanstående sannolikhet beräknas approximativt genom

$$P(X \leq 8) = P(X \leq 8.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{8.5 - np}{\sqrt{np(1-p)}}\right) \approx \Phi(-2.82) = 1 - \Phi(2.82) = 0.002.$$

även här ser man att H_0 förkastas på nivå 1%.

13.23 Låt x vara ett utfall av en $\text{Bin}(n, p)$ -fördelad stokastisk variabel X .

Hypotesen $H_0 : p = \frac{1}{2}$ förkastas till förmån för $H_1 : p \neq \frac{1}{2}$ då $|p_{\text{obs}}^* - 0.5| > 0.1$. Detta kan formuleras som att H_0 förkastas då

$$\left| \frac{x}{n} - \frac{1}{2} \right| > \frac{1}{10},$$

det vill säga då

$$x < \frac{4n}{10} \text{ eller } x > \frac{6n}{10}.$$

a) Med $n = 10$ fås att H_0 förkastas då $x < 4$ eller $x > 6$. Signifikansnivån blir då

$$\begin{aligned} P(\{X < 4\} \cup \{X > 6\}) &= 1 - \sum_{k=4}^6 P(X = k) = 1 - \sum_{k=4}^6 \binom{n}{k} p^k (1-p)^{n-k} = 1 - \sum_{k=4}^6 \binom{n}{k} \frac{1}{2^{10}} \\ &= 1 - \frac{210 + 252 + 210}{1024} = 0.3438. \end{aligned}$$

b) Med $n = 100$ fås att H_0 förkastas då $x < 40$ eller $x > 60$. Signifikansnivån blir då

$$P(\{X < 40\} \cup \{X > 60\}) = 1 - \sum_{k=40}^{60} P(X = k) = 0.0352.$$

Eftersom $np(1-p) = 100 \cdot 0.5 \cdot 0.5 = 25 > 10$ kan binomialfördelningen approximeras med normalfördelningen. Övanstående sannolikhet kan då beräknas

$$\begin{aligned} P(\{X < 40\} \cup \{X > 60\}) &= 1 - P(39.5 \leq X \leq 60.5) \\ &= 1 - P\left(\frac{39.5 - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{60.5 - np}{\sqrt{np(1-p)}}\right) \\ &\approx 1 - (\Phi(2.1) - \Phi(-2.1)) = 2(1 - \Phi(2.1)) = 0.0357. \end{aligned}$$

13.24 Antalet rätt utpekade vattenbehållare x är om personen chansar ett utfall av en hypergeometriskt fördelad variabel X , där $P(X = k) = \frac{\binom{5}{k} \binom{5}{5-k}}{\binom{10}{5}}$ för $k = 0, 1, \dots, 5$.

Hypotesen H_0 : slagrutan värdelös testas mot H_1 : slagrutan effektiv genom att H_0 förkastas för stora värden på x . Med utfallet $x = 4$ blir testets p-värde

$$P(X \geq 4) = P(X = 4) + P(X = 5) = \frac{\binom{5}{4} \binom{5}{1} + \binom{5}{5} \binom{5}{0}}{\binom{10}{5}} = \frac{25 + 1}{252} = 0.10317$$

det vill säga H_0 förkastas inte på nivå 10%.

13.25 Låt x_1, \dots, x_n vara observationer på oberoende $\text{Po}(\mu)$ -fördelade stokastiska variabler X_1, \dots, X_n .

Vi vill testa

$$H_0 : \mu = 0.2 = \mu_0 \quad \text{mot} \quad H_1 : \mu > \mu_0.$$

Vi förkastar H_0 för stora värden på $y = \sum_{i=1}^n x_i$ som om H_0 är sann är ett utfall på en $\text{Po}(n\mu_0) = \text{Po}(10)$ -fördelad stokastisk variabel Y . Vi observerar utfallet $y = 19$. Om H_0 är sann är

$$P(\text{Förkasta } H_0) = P(Y \geq 19) = 1 - P(Y \leq 18) = 1 - \sum_{k=0}^{18} \frac{10^k}{k!} e^{-10} = 1 - 0.99281 = 0.00719.$$

Detta är den lägsta signifikansnivå som H_0 kan förkastas på och vi ser att vi kan förkasta H_0 på nivå 1%.

13.26 Låt x_1, \dots, x_n vara utfall av oberoende $\text{Po}(\mu)$ -fördelade stokastiska variabler X_1, \dots, X_n .

Vi vill med $\mu_0 = 4$ testa

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu > \mu_0.$$

på signifikansnivå $\alpha = 0.001$ där sannolikheten för fel av andra slaget skall vara $\beta = 0.01$ då $\mu = 5$.

Antag att n är så stort att $Po(n\mu)$ -fördelningen kan approximeras med normalfördelningen.

Vi förkastar H_0 till förmån för H_1 för stora värden på z där

$$z = \frac{\sum_{i=1}^n x_i - n\mu_0}{\sqrt{n\mu_0}}$$

som då H_0 är sann är approximativt $N(0, 1)$. Ur $N(0, 1)$ -tabell fås att med $\lambda_{0.001} = 3.09$ och beslutsregeln att förkasta H_0 då $z > 3.09$ har testet signifikansnivå ≈ 0.001 .

Då $\mu = 5$ är styrkan

$$\begin{aligned} P(Z > \lambda_{0.001}) &= P\left(\frac{\sum_{i=1}^n X_i - n\mu_0}{\sqrt{n\mu_0}} > \lambda_{0.001}\right) = P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\mu}} > \frac{\lambda_{0.001}\sqrt{\mu_0} + \sqrt{n}(\mu_0 - \mu)}{\sqrt{\mu}}\right) \\ &\approx 1 - \Phi\left(\frac{\lambda_{0.001}\sqrt{\mu_0} + \sqrt{n}(\mu_0 - \mu)}{\sqrt{\mu}}\right). \end{aligned}$$

Kravet $P(Z > \lambda_{0.001}) \geq 0.99 = 1 - \Phi(-\lambda_{0.01})$ ger att

$$-\lambda_{0.01} \leq \frac{\lambda_{0.001}\sqrt{\mu_0} + \sqrt{n}(\mu_0 - \mu)}{\sqrt{\mu}}$$

eller

$$\sqrt{n} \geq \frac{\lambda_{0.01}\sqrt{\mu} + \lambda_{0.001}\sqrt{\mu_0}}{\mu - \mu_0} = \frac{2.3263\sqrt{5} + 3.090\sqrt{4}}{5 - 4} = 11.382$$

det vill säga $n \geq 129.56$. Alltså, tag $n = 130$ observationer och förkasta H_0 då

$$z = \frac{\sum_{i=1}^n x_i - n\mu_0}{\sqrt{n\mu_0}} = \frac{\sum_{i=1}^n x_i - 520}{\sqrt{520}}$$

är större än $\lambda_{0.001} = 3.09$.

Notera att $n = 130$ ger att $n\mu_0 = 520 > 15$ så normalapproximation av Poissonfördelningen är tillåten då H_0 är sann och även då styrkan för testet i $\mu = 5$ bestäms.

13.27 Låt x vara antalet par av $n = 18$ där det behandlade röret rostade mest. Vi antar modellen att x är ett utfall av den stokastiska variabeln X , där X är $\text{Bin}(n, p)$. Vi vill testa

$$H_0 : \text{behandlingen verkanlös} \quad \text{mot} \quad H_1 : \text{behandlingen effektiv.}$$

Hypoteserna formuleras i termer av modellens parameter som att vi vill testa

$$H_0 : p = \frac{1}{2} \quad \text{mot} \quad H_1 : p < \frac{1}{2}.$$

genom att förkasta H_0 då $x \leq 5$. Signifikansnivån (risknivån) är

$$\alpha = P(\text{förkasta korrekt } H_0) = P(X \leq 5) = \sum_{k=0}^5 \binom{n}{k} p^k (1-p)^{n-k} \Big|_{p=1/2} = 0.0481$$

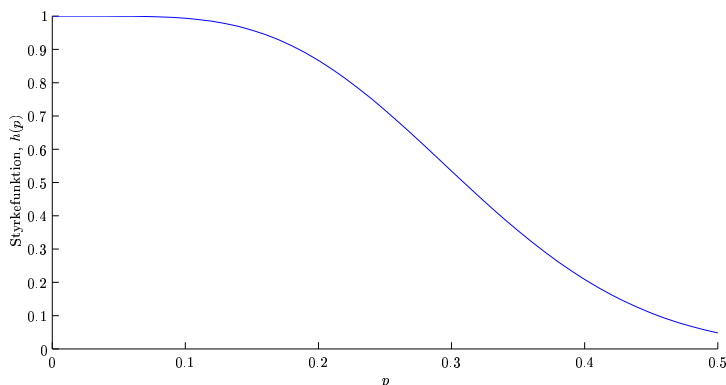
Slutsatsen om man ser utfallet...

- ... $x = 3 < 5$ är att vi förkastar H_0 till förmån för H_1 .
- ... $x = 7 \not< 5$ är att H_0 ej förkastas.
- ... $x = 15 \not< 5$ är att vi inte förkastar H_0 till förmån för H_1 .

Styrkan beräknas som

$$h(p) = P(\text{förfasta } H_0) = P(X \leq 5) = \sum_{k=0}^5 \binom{n}{k} p^k (1-p)^{n-k}$$

för $p \leq 1/2$. Då är $h(0.25) = 0.7175$.



Styrkefunktionen $h(p) = P(\text{förfasta } H_0)$ för $p \leq 1/2$. Ju längre från $H_0 : p = 1/2$ desto större är sannolikheten att nollhypotesen förfastas.

13.28 Låt x_1, \dots, x_r vara antalet marsvinsungar med färg motsvarande kategori $1, \dots, r$. Dessa modelleras som utfall av de (beroende) $\text{Bin}(n, p_i)$ -fördelade stokastiska variablerna X_1, \dots, X_r .

Enligt den genetiska modellen är p_i , sannolikheten att en unge får färg i , $9/16$, $3/16$ och $4/16$ för $i = 1, 2, 3$. Formellt, vi vill testa

$$H_0 : p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{4}{16}$$

på nivå $\alpha = 0.05$. Enligt modellen så är

	Kategori (färg)			
	1 (röd)	2 (svart)	3 (vit)	
Observerat antal: x_i	43	10	34	87
Hypotes: p_i	9/16	3/16	4/16	1
Förväntat antal: np_i	48.938	16.312	21.750	87

Vi förfastar H_0 för stora värden på

$$q = \sum_{i=1}^r \frac{(x_i - np_i)^2}{np_i}$$

som om H_0 är sann är ett utfall på en (approximativt) $\chi^2(r-1)$ -fördelad stokastisk variabel. Ur $\chi^2(3-1) = \chi^2(2)$ -tabeller får man att $\chi_{0.05}^2 = 5.99$. Med utfallet $q = 10.063 > \chi_{0.05}^2$ så förfastas H_0 på nivå 5%. (Vi förfastar även på 1%, testets p-värde är 0.65%.)

13.29 Låt x_1, \dots, x_r vara antalet utlånade böcker under de olika veckodagarna, med n som det totala antalet utlånade böcker under veckan,

$$n = \sum_{i=1}^r x_i = 623.$$

Notera att n egentligen borde modelleras som ett utfall av en stokastisk variabel, men vi kommer att räkna som om n vore fix. Då kan x_1, \dots, x_r modelleras som utfall av de (beroende) $\text{Bin}(n, p_i)$ -fördelade stokastiska variablerna X_1, \dots, X_r .

Vi vill testa en hypotes om att utlåningen är densamma oavsett veckodag mot att den varierar mellan veckodagar. Formellt: vi vill testa

$$H_0 : p_1 = p_2 = \dots = p_r = \frac{1}{r} \quad \text{mot} \quad H_1 : \text{inte } H_0$$

på signifikansnivå $\alpha = 0.05$. Enligt H_0 så är

	Kategori (veckodag)					
	1 (mån)	2 (tis)	3 (ons)	4 (tor)	5 (fre)	
Observerat antal: x_i	135	108	120	114	146	623
Hypotes: p_i	1/5	1/5	1/5	1/5	1/5	1
Förväntat antal: np_i	124.6	124.6	124.6	124.6	124.6	623

Vi förkastar H_0 för stora värden på

$$q = \sum_{i=1}^r \frac{(x_i - np_i)^2}{np_i}$$

som om H_0 är sann är ett utfall på en (approximativt) $\chi^2(r-1)$ -fördelad stokastisk variabel. Ur $\chi^2(5-1) = \chi^2(4)$ -tabeller får man att $\chi_{0.05}^2 = 9.49$. Med utfallet $q = 7.8266 < \chi_{0.05}^2$ så förkastas inte H_0 på nivå 5%.

På nivå $\alpha = 0.10$ får man ur $\chi^2(4)$ -tabeller att $\chi_{0.10}^2 = 7.78 < q$. Alltså förkastar vi H_0 på nivå 10%. (Testets p-värde är 9.81%.)

13.30 En stokastisk variabel X kan anta värdena $\Omega_X = \{0, 1, 2, 3\}$. Vi vill testa hypotesen

$$H_0 : X \text{ är Bin}(3, \frac{1}{4}) \quad \text{mot} \quad H_1 : X \text{ är ej Bin}(3, \frac{1}{4})$$

på nivå $\alpha = 0.01$ med hjälp av $n = 4096$ oberoende observationer av X . Med de möjliga värdena på X som kategorier är enligt H_0 sannolikheten att få en observation i kategori $i \in \Omega_X$

$$p_i = P(X = i) = \binom{3}{i} p^i (1-p)^{3-i} = \binom{3}{i} \left(\frac{1}{4}\right)^i \left(1 - \frac{1}{4}\right)^{3-i}, \quad i = 0, \dots, 3.$$

Låt x_i vara antalet utfall av X i kategori i , $i \in \Omega_X$. Om H_0 är sann så är

i	0	1	2	3	
observerad frekvens, x_i	1764	1692	552	88	4096
Hypotes, p_i	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$	1
Förväntat antal, np_i	1728	1728	576	64	4096

Vi förkastar H_0 för stora värden på

$$q = \sum_{i=1}^r \frac{(x_i - np_i)^2}{np_i}$$

som om H_0 är sann är ett utfall på en (approximativt) $\chi^2(r-1)$ -fördelad stokastisk variabel. Ur $\chi^2(4-1) = \chi^2(3)$ -tabeller får man att $\chi_{0.01}^2 = 11.35$. Med utfallet $q = 11.5 > \chi_{0.01}^2$ så förkastas H_0 på nivå 1%.

13.31 Låt x_{ij} vara antalet observationer i kategori j , $j = 1, \dots, r$, i population i , $i = 1, \dots, s$. Vi modellerar x_{ij} som ett utfall av en $\text{Bin}(n_i, p_j^{(i)})$ -fördelad stokastisk variabel X_{ij} .

x_{ij}	Kategori (kön)		
	1 (män)	2 (kvinnor)	
Population 1:	46	54	$100 = n_1$
Population 2:	78	72	$150 = n_2$
Population 3:	143	107	$250 = n_3$
Totalt:	$m_1 = 267$	$m_2 = 233$	$N = 500$

Vi vill testa om fördelningen av män och kvinnor skiljer sig mellan populationerna. Med H_0 som hypotesen att det inte finns någon skillnad,

$$H_0 : p_j^{(1)} = p_j^{(2)} = \dots = p_j^{(s)} \quad \text{för alla } j = 1, \dots, r,$$

så skattas p_j , sannolikheten att en utvald person har kön motsvarande kategori j , med

$$(p_j)_{\text{obs}}^* = \frac{1}{N} \sum_{i=1}^s x_{ij} = \frac{m_j}{N} \quad (p_1)_{\text{obs}}^* = \frac{267}{500}, \quad (p_2)_{\text{obs}}^* = \frac{233}{500}.$$

Vi kan då skatta det förväntade antalet observationer i kategori j i serie i , $E(X_{ij}) = n_i p_j$, med $n_i (p_j)_{\text{obs}}^* = \frac{n_i m_j}{N}$.

$\frac{n_i m_j}{N}$	Kategori (kön)		
	1 (män)	2 (kvinnor)	
Population 1:	53.4	46.6	100
Population 2:	80.1	69.9	150
Population 3:	133.5	116.5	250
Totalt:	267	233	500

Vi jämför de observerade antalen x_{ij} med de skattade förväntade antalen $\frac{n_i m_j}{N}$ med

$$q = \sum_{i=1}^s \sum_{j=1}^r \frac{(x_{ij} - \frac{n_i m_j}{N})^2}{\frac{n_i m_j}{N}}$$

som om H_0 är sann är ett utfall av en (approximativt) $\chi^2((s-1)(r-1))$ -fördelad stokastisk variabel. Vi förkastar H_0 för stora värden på q . Ur $\chi^2((3-1)(2-1)) = \chi^2(2)$ -tabeller får man att $\chi_{0.10}^2 = 4.61$. Vi observerar utfallet $q = 3.7694 < \chi_{0.10}^2$ så vi förkastar inte H_0 på nivå 10%. Det är på nivå 10% ingen signifikant skillnad i könsfördelning mellan populationerna.

13.32 De $n = 100$ observationerna på den förmodat ffg-fördelade stokastiska variabeln delas in i 9 kategorier:

Observerad frekvens: x_i	Kategori (värde)									100 = n
	1	2	3	4	5	6	7	8	≥ 9	
	42	23	10	11	8	2	3	1	0	

Notera att den sista kategorin är bara implicit given i uppgiften.

Låt p_i vara sannolikheten att en observation hamnar i kategori i enligt statistikerns modell. Om ffg(1/2)-antagandet stämmer är $p_i = (1 - 1/2)(1/2)^{i-1} = 2^{-i}$, $i < 9$, och $p_9 = 2^{-8}$, och det förväntade antalet observationer i kategori i , $E(X_i) = np_i$.

Observerad frekvens: x_i	Kategori (värde)									100
	1	2	3	4	5	6	7	8	≥ 9	
Hypotes: p_i	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	2^{-8}	1
Förväntat antal: np_i	50	25	12.5	6.25	3.125	1.5625	0.78125	0.39062	0.39062	100

Kategorier med ett lågt förväntat antal, $np_i < 5$, slås samman för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 5$ kategorier:

Observerad frekvens: x_i	Kategori (värde)					100
	1	2	3	4	≥ 5	
Hypotes: p_i	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-4}	1
Förväntat antal: np_i	50	25	12.5	6.25	6.25	100

Vi förkastar hypotesen om att observationerna kommer från en $\text{ffg}(1/2)$ -fördelning för stora värden på

$$q = \sum_{i=1}^r \frac{(x_i - np_i)^2}{np_i}$$

som då hypotesen stämmer är ett utfall från en (approximativt) $\chi^2(r-1)$ -fördelad stokastisk variabel. Ur $\chi^2(5-1) = \chi^2(4)$ -tabeller får man att $\chi_{0.05}^2 = 9.4877$. Vi observerar utfallet $q = 15.16 > \chi_{0.05}^2$ så vi förkastar hypotesen på nivå 5%.

På nivå 1% fås $\chi_{0.01}^2 = 13.277$ så hypotesen förkastas även på nivå 1%. (Testets p-värde är 0.438%.)

13.33 De $n = 300$ observationerna på den förmodat Poisson(μ)-fördelade stokastiska variabeln delas in i 4 kategorier beroende på dess värde:

	Kategori (värde)				
	0	1	2	≥ 3	
Observerad frekvens: x_i	249	42	9	0	300

Låt p_i vara sannolikheten att en observation hamnar i kategori i . Om Poisson(μ)-antagandet stämmer skattas μ med

$$\mu_{\text{obs}}^* = 0 \cdot \frac{249}{300} + 1 \cdot \frac{42}{300} + 2 \cdot \frac{9}{300} = 0.2$$

Med denna skattning erhålls skattningarna

$$(p_0)_{\text{obs}}^* = \frac{(\mu_{\text{obs}}^*)^0}{0!} e^{-\mu_{\text{obs}}^*} \quad (p_1)_{\text{obs}}^* = \frac{(\mu_{\text{obs}}^*)^1}{1!} e^{-\mu_{\text{obs}}^*} \quad (p_2)_{\text{obs}}^* = \frac{(\mu_{\text{obs}}^*)^2}{2!} e^{-\mu_{\text{obs}}^*} \quad (p_3)_{\text{obs}}^* = 1 - \sum_{i=0}^2 (p_i)_{\text{obs}}^*.$$

De (skattade) förväntade antalet observationer i kategori i är $n(p_i)_{\text{obs}}^*$.

	Kategori (värde)				
	0	1	2	≥ 3	
Observerad frekvens: x_i	249	42	9	0	$300 = n$
Hypotes: $(p_i)_{\text{obs}}^*$	0.81873	0.16375	0.016375	0.0011485	1
Förväntat antal: $n(p_i)_{\text{obs}}^*$	245.62	49.124	4.9124	0.34454	300

Kategorier med ett lågt (skattat) förväntat antal, $n(p_i)_{\text{obs}}^*$, slås samman för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 3$ kategorier:

	Kategori (värde)			
	0	1	≥ 2	
Observerad frekvens: x_i	249	42	9	$300 = n$
Hypotes: $(p_i)_{\text{obs}}^*$	0.81873	0.16375	0.017523	1
Förväntat antal: $n(p_i)_{\text{obs}}^*$	245.62	49.124	5.2569	300

Vi förkastar en hypotes om Poisson(μ)-fördelning för stora värden på

$$q = \sum_{i=1}^r \frac{(x_i - n(p_i)_{\text{obs}}^*)^2}{n(p_i)_{\text{obs}}^*}$$

som då hypotesen stämmer är ett utfall av en (approximativt) $\chi^2(r-1-1)$ -fördelad stokastisk variabel. En extra frihetsgrad har försvunnit då parametern μ ersattes med skattningen μ_{obs}^* . Ur $\chi^2(3-1-1) = \chi^2(1)$ -tabeller får man att $\chi_{0.05}^2 = 3.8415$. Vi observerar utfallet $q = 3.7448 < 3.8415$ så vi förkastar inte hypotesen på nivå 5%. Vi kan inte utesluta att observationerna kommer från en Poissonfördelning.

På nivå 10% fås $\chi_{0.10}^2 = 2.7055 < q$ så hypotesen förkastas på nivå 10%. (Testets p-värde är 5.30%.)

13.34 De $n = 81$ observationerna på den förmodat Poisson-fördelade stokastiska variabeln delas in i 8 kategorier beroende på dess värde:

	Kategori (antal bilar)								
	0	1	2	3	4	5	6	≥ 7	
Observerad frekvens: x_i	14	12	25	16	10	3	1	0	$81 = n$

Notera att den sista kategorin är bara implicit given i uppgiften och att n egentligen borde modelleras som ett utfall av en stokastisk variabel, men vi kommer att räkna som om n vore fix.

Låt p_i vara sannolikheten att antalet bilar motsvarande kategori i passerar under tidsintervallet. Om antalet bilar beskrivs av en Poisson(μ)-fördelning så skattas μ med

$$\mu_{\text{obs}}^* = 0 \cdot \frac{14}{81} + 1 \cdot \frac{12}{81} + \dots + 6 \cdot \frac{1}{81} + 7 \cdot \frac{0}{81} = 2.1111.$$

Med denna skattning kan

$$p_i = \frac{\mu^i}{i!} e^{-\mu} \quad \text{skattas med} \quad (p_i)_{\text{obs}}^* = \frac{(\mu_{\text{obs}}^*)^i}{i!} e^{-\mu_{\text{obs}}^*}$$

för $i < 7$ och $(p_7)_{\text{obs}}^* = 1 - \sum_{i=0}^6 (p_i)_{\text{obs}}^*$. De (skattade) förväntade antalet observationer i kategori i är $n(p_i)_{\text{obs}}^*$.

	Kategori (antal bilar)								
	0	1	2	3	4	5	6	≥ 7	
Observerad frekvens: x_i	14	12	25	16	10	3	1	0	81
Hypotes: $(p_i)_{\text{obs}}^*$	0.1211	0.25566	0.26987	0.18991	0.10023	0.042319	0.01489	0.0060259	1
Förväntat antal: $n(p_i)_{\text{obs}}^*$	9.8094	20.709	21.859	15.382	8.1185	3.4278	1.2061	0.4881	81

Kategorier med ett lågt förväntat antal, $n(p_i)_{\text{obs}}^* < 5$, slås samman för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 6$ kategorier:

	Kategori (antal bilar)						
	0	1	2	3	4	≥ 5	
Observerad frekvens: x_i	14	12	25	16	10	4	81
Hypotes: $(p_i)_{\text{obs}}^*$	0.1211	0.25566	0.26987	0.18991	0.10023	0.063234	1
Förväntat antal: $n(p_i)_{\text{obs}}^*$	9.8094	20.709	21.859	15.382	8.1185	5.122	81

Vi jämför de observerade antalen x_i med de skattade förväntade antalen $n(p_i)_{\text{obs}}^*$ med teststatistikan

$$q = \sum_{i=1}^r \frac{(x_i - n(p_i)_{\text{obs}}^*)^2}{n(p_i)_{\text{obs}}^*} = 6.61$$

som då hypotesen om Poisson-fördelning stämmer är ett utfall av en (approximativt) $\chi^2(r - 1 - 1)$ -fördelad stokastisk variabel. Notera att en extra frihetsgrad försvinner då parametern μ skattas. Hypotesen förkastas för stora värden på q och ur $\chi^2(6 - 1 - 1) = \chi^2(4)$ -tabeller får man att $\chi_{0.05}^2 = 9.49$. Här är $q = 6.61 < 9.49$ så vi förkastar ej hypotesen om Poissonfördelning på risknivå 5%. Vi kan inte utesluta att observationerna kommer från en Poissonfördelning.

13.35 Låt x_{ij} vara antalet observationer i kategori (i, j) , $j = 1, \dots, r$, $i = 1, \dots, s$. Beteckna det totala antalet skadade med $N = \sum_{i=1}^s \sum_{j=1}^r x_{ij} = 500$. Notera att N egentligen borde modelleras som ett utfall av en stokastisk variabel, men vi kommer att räkna som om N vore fix. Då kan x_{ij} modelleras som utfall av en

Bin(N, p_{ij})-fördelad stokastisk variabel X_{ij} .

x_{ij}		Säkerhetsbälte		
		1 (användes)	2 (användes ej)	
Personskador	1 (lätta)	101	143	$n_1 = 244$
	2 (svåra)	58	198	$n_2 = 256$
		$m_1 = 159$	$m_2 = 341$	$N = 500$

Vi skattar fördelningen för personskador med

$$(p_i)_{\text{obs}}^* = \frac{n_i}{N} \quad (p_1)_{\text{obs}}^* = \frac{244}{500}, \quad (p_2)_{\text{obs}}^* = \frac{256}{500}.$$

Vi skattar fördelningen för användandet av säkerhetsbälte med

$$(q_i)_{\text{obs}}^* = \frac{m_i}{N} \quad (q_1)_{\text{obs}}^* = \frac{159}{500}, \quad (q_2)_{\text{obs}}^* = \frac{341}{500}.$$

Om graden av personskada är oberoende av användandet av säkerhetsbälte, $p_{ij} = p_i q_j$, kan p_{ij} skattas med $(p_i)_{\text{obs}}^* \cdot (q_j)_{\text{obs}}^*$ och det förväntade antalet observationer med

$$N (p_i)_{\text{obs}}^* \cdot (q_j)_{\text{obs}}^* = \frac{n_i m_j}{N}.$$

$\frac{n_i m_j}{N}$		Säkerhetsbälte		
		1 (användes)	2 (användes ej)	
Personskador	1 (lätta)	77.592	166.408	244
	2 (svåra)	81.408	174.592	256
		159	341	500

Vi förkastar en hypotes om oberoende mellan grad av personskada och användande av säkerhetsbälte för stora värden på

$$q = \sum_{i=1}^s \sum_{j=1}^r \frac{(x_{ij} - \frac{n_i m_j}{N})^2}{\frac{n_i m_j}{N}}$$

som om hypotesen stämmer är ett utfall av en (approximativt) $\chi^2((s-1)(r-1))$ -fördelad stokastisk variabel. Ur $\chi^2((2-1)(2-1)) = \chi^2(1)$ -tabeller får man att $\chi_{0.01}^2 = 6.63$. Vi observerar utfallet $q = 20.2235 > \chi_{0.01}^2$ så vi förkastar hypotesen om oberoende på nivå 1%. Användandet av säkerhetsbälte och graden av personskada är inte oberoende. (Testets p-värde är $\approx 6.9 \cdot 10^{-6}$.)

14.1 Låt y_1, \dots, y_n vara de uppmätta ljusextinktionerna vid koncentrationerna x_1, \dots, x_n . Vi antar modellen att y_1, \dots, y_n är utfall av oberoende $N(\alpha + \beta x_i, \sigma)$ -fördelade stokastiska variabler Y_1, \dots, Y_n .

Koncentration, x_1, \dots, x_n	0.40	0.70	1.00	1.20	1.40	1.70	2.00
Extinktion, y_1, \dots, y_n	0.23	0.34	0.42	0.55	0.61	0.77	0.84.

Observationerna sammanfattas av storheterna

$$\bar{x} = 1.20, \quad \bar{y} = 0.537$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 1.86, \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 0.741, \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = 0.298.$$

Parametrarna β och α skattas med

$$\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} = 0.398 \quad \text{respektive} \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = 0.0591$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2} [S_{yy} - \beta_{\text{obs}}^* S_{xy}] = 0.0013888,$$

det vill säga σ skattas med $s = 0.025052$.

a) Per enhet ökar extinktionen med faktorn β som skattas med $\beta_{\text{obs}}^* = 0.398$.

b) En skattning av $E(Y(1.20)) = \alpha + \beta \cdot 1.20$ är $\alpha_{\text{obs}}^* + \beta_{\text{obs}}^* \cdot 1.20 = 0.53714$. Skattningen $\alpha_{\text{obs}}^* + \beta_{\text{obs}}^* \cdot 1.20$ är ett utfall av en stokastisk variabel med varians

$$\begin{aligned} V(\alpha^* + \beta^* \cdot 1.2) &= V(\bar{Y} - \beta^* \bar{x} + \beta^* \cdot 1.2) = V(\bar{Y} + \beta^*(1.2 - \bar{x})) = \{\text{oberoende}\} \\ &= V(\bar{Y}) + (1.2 - \bar{x})^2 V(\beta^*) = \frac{\sigma^2}{n} + (1.2 - \bar{x})^2 \frac{\sigma^2}{S_{xx}}. \end{aligned}$$

Alltså, standardavvikelsen är

$$D(\alpha^* + \beta^* \cdot 1.2) = \sigma \sqrt{\frac{1}{n} + \frac{(1.2 - \bar{x})^2}{S_{xx}}} = \sigma \sqrt{\frac{1}{7} + \frac{(1.2 - 1.2)^2}{S_{xx}}} = \sigma / \sqrt{7}.$$

c) En skattning av $E(Y(0)) = \alpha$ är $\alpha_{\text{obs}}^* = 0.0591$. Skattningen α_{obs}^* är ett utfall av en stokastisk variabel med varians

$$V(\alpha^*) = V(\bar{Y} - \beta^* \bar{x}) = \{\text{oberoende}\} = V(\bar{Y}) + (-\bar{x})^2 V(\beta^*) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{S_{xx}}.$$

Alltså, standardavvikelsen är

$$D(\alpha^*) = \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} = \sigma \sqrt{\frac{1}{7} + \frac{(1.2)^2}{1.86}} = 0.9576\sigma.$$

d) En skattning av σ är $s = 0.025052$.

14.2 Låt y_1, \dots, y_n vara sådana att $y_i = \alpha + \beta x_i + z_i$ där $\alpha, \beta, x_1, \dots, x_n$ är konstanter och z_1, \dots, z_n är utfall av oberoende $N(0, \sigma)$ -fördelade stokastiska variabler. Då är y_1, \dots, y_n utfall av oberoende $N(\alpha + \beta x_i, \sigma)$ -fördelade stokastiska variabler. Med $n = 10$ observationer erhålls

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.20, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 1.50$$

och

$$\begin{aligned} S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = 18.40 - 10(1.20)^2 = 4.00 \\ S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = 20.40 - 10(1.20)(1.50) = 2.40 \\ S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = 27.86 - 10(1.50)^2 = 5.36. \end{aligned}$$

Parametrarna β och α skattas med

$$\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} = 0.60 \quad \text{respektive} \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = 0.78.$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2} [S_{yy} - \beta_{\text{obs}}^* S_{xy}] = 0.49,$$

det vill säga σ skattas med $s = 0.70$.

Skattningen β_{obs}^* beskrivs av β^* där β^* är $N(\beta, \sigma/\sqrt{S_{xx}})$. Ur $t(n-2) = t(8)$ -tabeller får man att $t_{0.025} = 2.31$ och ett 95% konfidensintervall för β ges av

$$\beta \in \beta_{\text{obs}}^* \pm t_{0.025} \frac{s}{\sqrt{S_{xx}}} = 0.60 \pm 2.31 \cdot \frac{0.70}{\sqrt{4.00}} = 0.60 \pm 0.81 = (-0.21, 1.41).$$

Eftersom α^* är $N\left(\alpha, \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right)$ så ges ett 95% konfidensintervall för α av

$$\alpha \in \alpha_{\text{obs}}^* \pm t_{0.025} s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} = 0.78 \pm 2.31 \cdot 0.70 \cdot \sqrt{\frac{1}{10} + \frac{(1.20)^2}{4.00}} = 0.78 \pm 1.10 = (-0.32, 1.88).$$

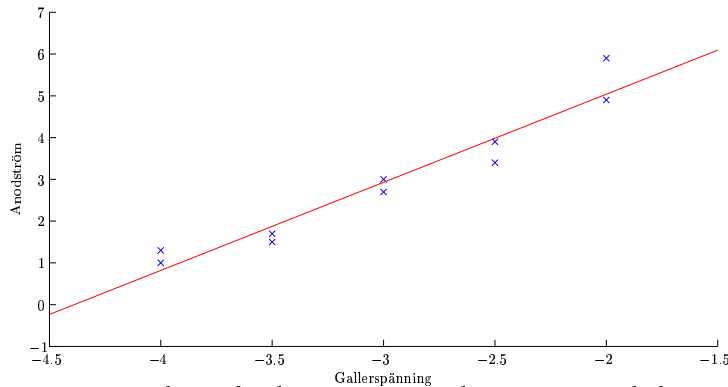
14.3 Låt y_1, \dots, y_n vara de uppmätta anodströmmarna vid gallerspänningarna x_1, \dots, x_n . Vi ansätter modellen att y_1, \dots, y_n är utfall av oberoende $N(\alpha + \beta x_i, \sigma)$ -fördelade stokastiska variabler Y_1, \dots, Y_n . De $n = 10$ observationerna sammanfattas av

$$\bar{x} = -3.00, \quad \bar{y} = 2.93$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 5.00, \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 10.55, \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = 23.861.$$

Parametrarna β och α skattas med

$$\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} = 2.11 \quad \text{respektive} \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = 9.26.$$



Regressionslinjen för de uppmätta anodströmmarna vid olika gallerspänningar.

Variansen σ^2 skattas med

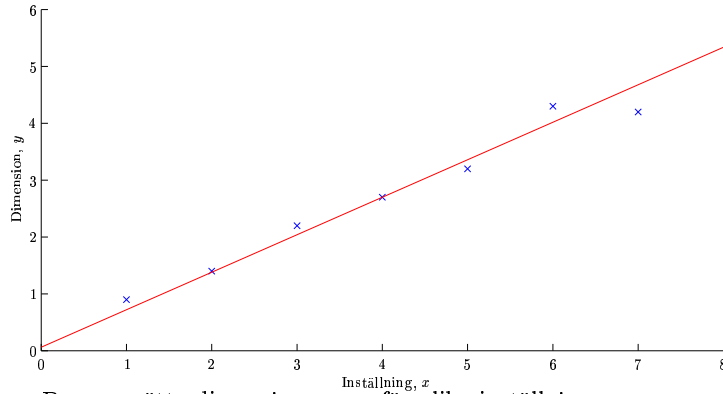
$$s^2 = \frac{1}{n-2} [S_{yy} - \beta_{\text{obs}}^* S_{xy}] = 0.200.$$

Skattningen β_{obs}^* beskrivs av β^* där β^* är $N(\beta, \sigma/\sqrt{S_{xx}})$. Ur $t(n-2) = t(8)$ -tabeller får man att $t_{0.025} = 2.31$ och ett 95% konfidensintervall för β ges av

$$\beta \in \beta_{\text{obs}}^* \pm t_{0.025} \frac{s}{\sqrt{S_{xx}}} = 2.11 \pm 2.31 \cdot \frac{\sqrt{0.200}}{\sqrt{5.00}} = 2.11 \pm 0.46 = (1.65, 2.57).$$

14.4 Låt y_1, \dots, y_n vara de uppmätta dimensionerna vid inställningarna x_1, \dots, x_n . Vi ansätter modellen att y_1, \dots, y_n är utfall av oberoende $N(\alpha + \beta x_i, \sigma)$ -fördelade stokastiska variabler Y_1, \dots, Y_n .

Inställning, x_1, \dots, x_n	1.0	2.0	3.0	4.0	5.0	6.0	7.0
Dimension, y_1, \dots, y_n	0.9	1.4	2.2	2.7	3.2	4.3	4.2



De uppmätta dimensionerna y för olika inställningar x markerade med kryss med den skattade regressionslinjen inritad.

Observationerna sammanfattas av storheterna

$$\bar{x} = 4.0, \quad \bar{y} = 2.7$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 28, \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 16.7, \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = 10.24.$$

b) Parametrarna β och α skattas med

$$\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} = 0.59643 \quad \text{respektive} \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = 0.31429.$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2} [S_{yy} - \beta_{\text{obs}}^* S_{xy}] = 0.055929,$$

det vill säga σ skattas med $s = 0.23649$.

c) Om man önskar att $2.5 = E(Y(x)) = \alpha + \beta x$ skall x väljas så att $x = (2.5 - \alpha)/\beta$. Denna storhet skattas med $(2.5 - \alpha_{\text{obs}}^*)/\beta_{\text{obs}}^* = 3.6647$.

d) Skattningen α_{obs}^* är ett utfall av den normalfördelade stokastiska variabeln α^* , där

$$\alpha^* \quad \text{är} \quad N \left(\alpha, \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \right).$$

Ur $t(n-2) = t(7-2)$ -tabell erhålles att $t_{0.025} = 2.57$. Alltså ges ett konfidensintervall för α av

$$\alpha \in \alpha_{\text{obs}}^* \pm t_{0.025} s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} = 0.31429 \pm 2.57 \cdot 0.19987 = 0.314 \pm 0.514 \quad (95\%).$$

Skattningen β_{obs}^* är ett utfall av den normalfördelade stokastiska variabeln β^* , där

$$\beta^* \quad \text{är} \quad N \left(\beta, \frac{\sigma}{\sqrt{S_{xx}}} \right).$$

Alltså ges ett konfidensintervall för β av

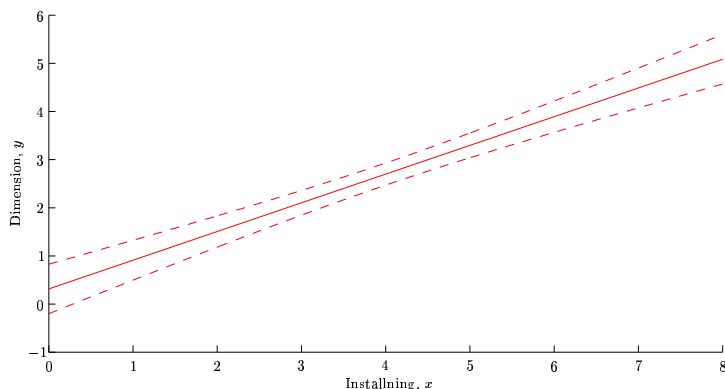
$$\beta \in \beta_{\text{obs}}^* \pm t_{0.025} \frac{s}{\sqrt{S_{xx}}} = 0.59643 \pm 2.57 \cdot 0.044693 = 0.596 \pm 0.115 \quad (95\%).$$

e) För en punkt x_0 skattas $E(Y(x_0)) = \alpha + \beta x_0$ av $\alpha_{\text{obs}}^* + \beta_{\text{obs}}^* x_0$ som är ett utfall av $\alpha^* + \beta^* x_0$, där

$$\alpha^* + \beta^* x_0 \quad \text{är} \quad N \left(\alpha + \beta x_0, \sigma \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right).$$

Alltså ges ett konfidensintervall för $\alpha + \beta x_0$ av

$$\alpha + \beta x_0 \in \alpha_{\text{obs}}^* + \beta_{\text{obs}}^* x_0 \pm t_{0.025, s} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$



Regressionslinjen (heldragen) med konfidensintervallet för väntevärdet $E(Y(x))$ som funktion av x (streckade linjer).

14.5 Vi gör mätningar i punkter x svarande mot $x = 1/\lambda^2$ och får följande observationer:

λ_i	6232	5571	5100
$x_i = 1/\lambda_i^2$	$2.5748 \cdot 10^{-8}$	$3.2221 \cdot 10^{-8}$	$3.8447 \cdot 10^{-8}$
y_i	19.38	25.62	30.10

Data kan sammanfattas med

$$\bar{x} = 3.2138 \cdot 10^{-8} \quad \bar{y} = 25.033$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = 8.0638 \cdot 10^{-17} \quad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 6.8137 \cdot 10^{-8} \quad S_{yy} = \sum (y_i - \bar{y})^2 = 57.975.$$

Vi ansätter modellen att y_1, \dots, y_n är utfall av oberoende $N(\alpha + \beta x_i, \sigma)$ -fördelade stokastiska variabler Y_1, \dots, Y_n .

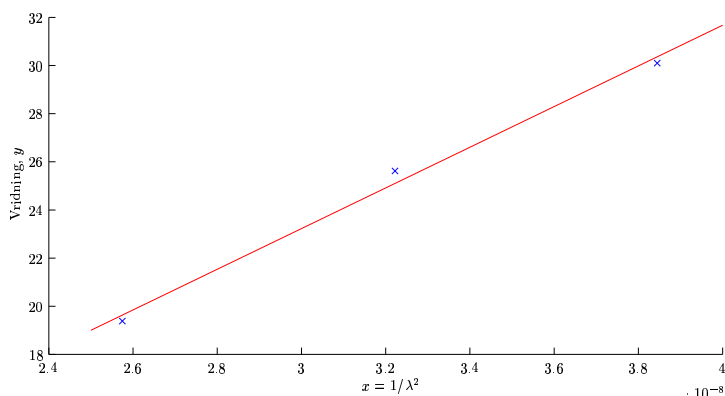
a) Parametrarna β och α skattas med

$$\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} = 8.45 \cdot 10^8 \quad \text{respektive} \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = -2.123.$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2} [S_{yy} - \beta_{\text{obs}}^* S_{xy}] = 0.40141,$$

det vill säga σ skattas med $s = 0.63357$.



De tre mätningarna på vridningen som funktion av $1/\lambda^2$, där λ är ljusets våglängd, med inritad regressionslinje.

c) Skattningen β_{obs}^* är ett utfall av den normalfördelade stokastiska variabeln β^* , där

$$\beta^* \text{ är } N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right).$$

Ur $t(n-2) = t(3-2)$ -tabell erhålles att $t_{0.025} = 12.706$ (!). Alltså ges ett konfidensintervall för β av

$$\beta \in \beta_{\text{obs}}^* \pm t_{0.025} \frac{s}{\sqrt{S_{xx}}} = 8.45 \cdot 10^8 \pm 12.706 \cdot 7.0555 \cdot 10^7 = (8.45 \pm 8.96) \cdot 10^8 \quad (95\%).$$

14.6 Korrelationskoefficienten kan med standardbeteckningar skrivas som

$$\begin{aligned} r &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{j=1}^n (y_j - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \\ &= \text{sgn}(S_{xy}) \cdot \sqrt{\frac{S_{xy}}{S_{xx}} \cdot \frac{S_{xy}}{S_{yy}}}, \end{aligned}$$

där funktionen $\text{sgn}(S_{xy}) = S_{xy}/|S_{xy}|$ om $|S_{xy}| \neq 0$, 0 annars. Regressionslinjernas parametrar

$$y(x) = (\alpha_x)_{\text{obs}}^* + (\beta_x)_{\text{obs}}^* x = 6.06 + (-2.43)x \quad \text{och} \quad x(y) = (\alpha_y)_{\text{obs}}^* + (\beta_y)_{\text{obs}}^* y = 1.30 + (-0.31)y$$

är beräknade med

$$(\beta_x)_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} \quad \text{och} \quad (\beta_y)_{\text{obs}}^* = \frac{S_{xy}}{S_{yy}}$$

så

$$r = \text{sgn}(S_{xy}) \sqrt{\frac{S_{xy}}{S_{xx}} \cdot \frac{S_{xy}}{S_{yy}}} = (-1) \sqrt{(\beta_x)_{\text{obs}}^* (\beta_y)_{\text{obs}}^*} = -\sqrt{(-2.43)(-0.31)} = -0.868.$$

Vidare så är

$$\begin{cases} (\alpha_x)_{\text{obs}}^* &= \bar{y} - (\beta_x)_{\text{obs}}^* \bar{x} \\ (\alpha_y)_{\text{obs}}^* &= \bar{x} - (\beta_y)_{\text{obs}}^* \bar{y} \end{cases} \quad \text{det vill säga} \quad \begin{cases} 6.06 &= \bar{y} + 2.43 \bar{x} \\ 1.30 &= \bar{x} + 0.31 \bar{y} \end{cases}$$

vilket har lösningen $\bar{x} = -2.3454$ och $\bar{y} = 11.756$.

14.7 a) Om hypotesen $\beta_1 = 0$ förkastas till förmån för $\beta_1 \neq 0$ har testet p-värde $8.3 \cdot 10^{-6}$. Detta är den lägsta signifikansnivå hypotesen kan förkastas på och speciellt förkastas den på nivå 5%.

Om hypotesen $\beta_2 = 0$ förkastas till förmån för $\beta_2 \neq 0$ har testet p-värde 0.132. Detta är den lägsta signifikansnivå hypotesen kan förkastas på. Alltså, hypotesen förkastas ej på nivå 5%.

b) Förändringen $\beta_1 + \beta_2$ skattas med $(\beta_1)_{\text{obs}}^* + (\beta_2)_{\text{obs}}^* = 1.423 - 0.176 = 1.247$.

c) Skattningen $(\beta_1)_{\text{obs}}^* + (\beta_2)_{\text{obs}}^*$ är ett utfall av en stokastisk variabel $\beta_1^* + \beta_2^*$ med varians

$$V(\beta_1^* + \beta_2^*) = V(\beta_1^*) + V(\beta_2^*) + 2C(\beta_1^*, \beta_2^*),$$

det vill säga standardavvikelsen skattas till

$$d(\beta_1^* + \beta_2^*) = \sqrt{(0.247)^2 + (0.112)^2 + 2 \cdot (-0.00248)} = 0.2619.$$

14.8 Vi ansätter modellen att Y_0, \dots, Y_n är oberoende $N(v_i t_i + at_i^2, \sigma)$ -fördelade stokastiska variabler där $v_1 = v_2 = \dots = v_n = 0$ och $\sigma = 0.070$. Notera att denna a -parameter är hälften så stor som parametern i uppgiftens formulering (accelerationen). Minsta-kvadratskattningarna av a och v_0 är de värden på a och v_0 som minimerar

$$Q(a, v_0) = \sum_{i=0}^n (y_i - E(Y_i))^2 = (y_0 - (v_0 t_0 + at_0^2))^2 + \sum_{i=1}^n (y_i - at_i^2)^2.$$

Lösning av ekvationssystemet

$$\begin{cases} 0 = \frac{\partial}{\partial a} Q(a, v_0) &= (-2)(y_0 - (v_0 t_0 + at_0^2))t_0^2 + (-2) \sum_{i=1}^n (y_i - at_i^2)t_i^2 \\ &= (-2) \left(\sum_{i=0}^n y_i t_i^2 - v_0 t_0^3 - a \sum_{i=0}^n t_i^4 \right) \\ 0 = \frac{\partial}{\partial v_0} Q(a, v_0) &= (-2)(y_0 - (v_0 t_0 + at_0^2))t_0 = (-2)t_0 (y_0 - v_0 t_0 - at_0^2) \end{cases}$$

ger

$$\begin{cases} v_0 &= \frac{y_0}{t_0} - at_0 = \frac{y_0}{t_0} - t_0 \frac{\sum_{i=1}^n y_i t_i^2}{\sum_{i=1}^n t_i^4} \\ a &= \frac{\sum_{i=1}^n y_i t_i^2}{\sum_{i=1}^n t_i^4} \end{cases}$$

Alltså, skattningen blir

$$(v_0)_{\text{obs}}^* = \frac{y_0}{t_0} - t_0 \frac{\sum_{i=1}^n y_i t_i^2}{\sum_{i=1}^n t_i^4} = 0.31255.$$

Detta är ett utfall av den stokastiska variabeln

$$v_0^* = \frac{Y_0}{t_0} - t_0 \frac{\sum_{i=1}^n Y_i t_i^2}{\sum_{i=1}^n t_i^4}$$

som har väntevärde

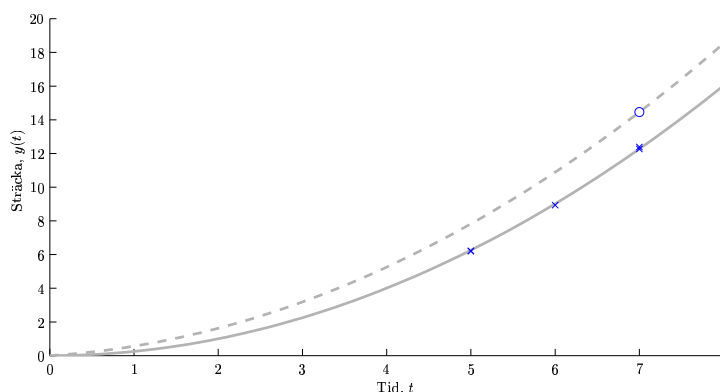
$$E(v_0^*) = E\left(\frac{Y_0}{t_0} - t_0 \frac{\sum_{i=1}^n Y_i t_i^2}{\sum_{i=1}^n t_i^4}\right) = \frac{v_0 t_0 + at_0^2}{t_0} - t_0 \frac{\sum_{i=1}^n at_i^2 \cdot t_i^2}{\sum_{i=1}^n t_i^4} = (v_0 + at_0) - at_0 = v_0$$

och varians

$$V(v_0^*) = V\left(\frac{Y_0}{t_0} - t_0 \frac{\sum_{i=1}^n Y_i t_i^2}{\sum_{i=1}^n t_i^4}\right) = \frac{1}{t_0^2} \sigma^2 + \frac{t_0^2}{(\sum_{i=1}^n t_i^4)^2} \sum_{i=1}^n (t_i^2)^2 \sigma^2 = \sigma^2 \left(\frac{1}{t_0^2} + \frac{t_0^2}{\sum_{i=1}^n t_i^4} \right).$$

Eftersom v_0^* är en linjärkombination av oberoende normalfördelade stokastiska variabler är v_0^* normalfördelad med väntevärde v_0 och varians enligt ovan. Ett konfidensintervall för v_0 med konfidensgrad 95% ges sedan av

$$v_0 \in (v_0)_{\text{obs}}^* \pm \lambda_{0.025} \sigma \sqrt{\frac{1}{t_0^2} + \frac{t_0^2}{\sum_{i=1}^n t_i^4}} = 0.313 \pm 1.96 \cdot 0.0115 = 0.313 \pm 0.023 = (0.290, 0.335).$$



Regressionskurvan motsvarande väntevärdet $at^2 = E(Y_1(t))$ (heldragen) och regressionskurvan motsvarande väntevärdet $v_0t + at^2 = E(Y_0(t))$ (streckad).

På matrisform kan man lösa problemet såsom följer. Låt

$$\boldsymbol{\beta} = \begin{bmatrix} v_0 \\ a \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} t_0 & t_0^2 \\ 0 & t_1^2 \\ 0 & t_2^2 \\ 0 & t_3^2 \\ 0 & t_4^2 \\ 0 & t_5^2 \end{bmatrix} = \begin{bmatrix} 7 & 49 \\ 0 & 25 \\ 0 & 25 \\ 0 & 36 \\ 0 & 49 \\ 0 & 49 \end{bmatrix} \quad \text{och} \quad \mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 14.46 \\ 6.22 \\ 6.20 \\ 8.93 \\ 12.27 \\ 12.39 \end{bmatrix}.$$

Vi söker minsta-kvadratlösningen till $\mathbf{A}\boldsymbol{\beta} = \mathbf{y}$. Genom att vänstermultiplicera med \mathbf{A}^T fås ekvationssystemet $\mathbf{A}^T\mathbf{A}\boldsymbol{\beta} = \mathbf{A}^T\mathbf{y}$ vilket har lösningen

$$\boldsymbol{\beta}_{\text{obs}}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \begin{bmatrix} 0.31255 \\ 0.25045 \end{bmatrix}$$

varur $(v_0)_{\text{obs}}^* = 0.31255$ och $a_{\text{obs}}^* = 0.25045$ utläses. Kovariansmatrisen

$$\begin{aligned} C(\boldsymbol{\beta}^*, \boldsymbol{\beta}^*) &= C\left((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}, (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}\right) = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T C(\mathbf{Y}, \mathbf{Y}) ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)^T \\ &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\sigma^2\mathbf{I}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-T} = \sigma^2(\mathbf{A}^T\mathbf{A})^{-T} = \begin{bmatrix} 0.00013268 & -4.6679 \cdot 10^{-6} \\ -4.6679 \cdot 10^{-6} & 6.6685 \cdot 10^{-7} \end{bmatrix}, \end{aligned}$$

så $V(v_0^*) = 0.00013268$ och $D(v_0^*) = 0.011518$. Konfidensintervall erhålls på samma sätt som tidigare.