LÖSNINGSFÖRSLAG TILL UPPGIFTER I PROBLEMSAMLINGEN I MATEMATISK STATISTIK

Version 9 december 2004

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 nogrannhet.

2.1

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 3 element

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

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

c) (1) Utfallsrummet är överuppräkneligt och ges av

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

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

(2) Utfallsrummet är överuppräkneligt 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 överuppräkneligt och ges av

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

2.2 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 $A = \{(x, y) \in \Omega : x + y < 6\}$, dvs. 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 $P(A) = 10/36$.

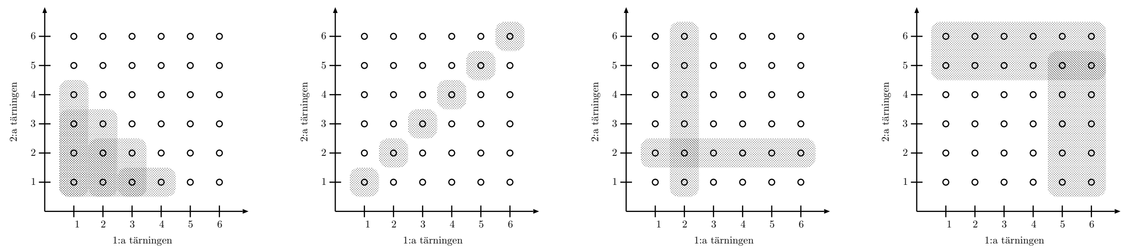
b) $B = \{\text{Samma poäng vid båda kasten}\}$ så är $B = \{(x, x) \in \Omega\}$, dvs. B består av de 6 utfallen $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$. Således $P(B) = 1/6$. Man kan också tänka sig att man kastar den ena tärningen före den andra. Den första tärningen bestämmer det värde som man skall träffa med den andra tärningen. För varje utfall på den första tärningen är sannolikheten 1/6 att tärning 2 kommer att visa samma värde (tärningarna är oberoende).

c)

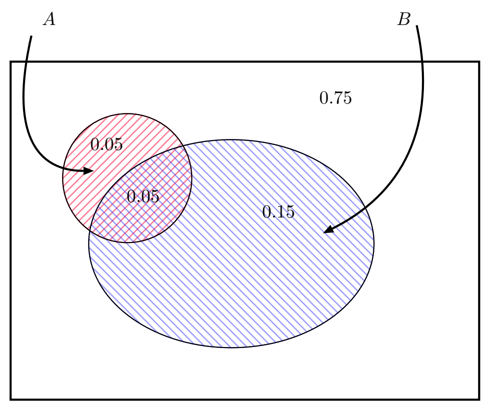
$$\begin{aligned} C &= \left\{\text{Åtminstone ett av kasten ger precis två poäng}\right\} \\ &= \{(1, 2), (2, 2), (3, 2), (4, 2), (5, 2), (6, 2), (2, 1), (2, 3), (2, 4), (2, 5), (2, 6)\} \end{aligned}$$

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

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



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

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

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

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

2.4 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, A och B kan inte vara oförenliga (disjunkta).

2.5 Rita figur!

a) Eftersom A och B är disjunkta så är $P(A \cup B) = P(A) + P(B) = 0.3 + 0.5 = 0.8.$

b) Additionsformeln för unioner ger $P(A \cup C) = P(A) + P(C) - P(A \cap C) = 0.3 + 0.7 - 0.2 = 0.8.$

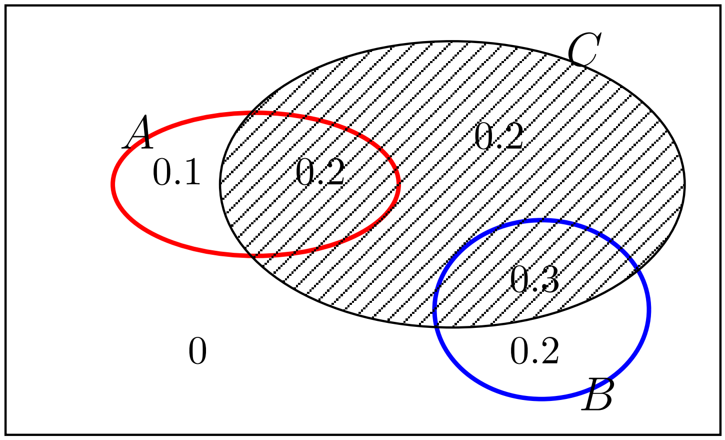
c) Additionsformeln för unioner ger $P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0.5 + 0.7 - 0.3 = 0.9.$

d) Slutligen,

$$\begin{aligned} P((A \cup B) \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - (P(A \cap C) + P(B \cap C)) \\ &= 0.3 + 0.5 + 0.7 - (0.2 + 0.3) = 1. \end{aligned}$$

alternativt kan sannolikheten beräknas via additionsformeln för tre händelser

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= 0.3 + 0.5 + 0.7 - 0 - 0.2 - 0.3 + 0 = 1. \end{aligned}$$



Venndiagram där A och B är disjunkta.

2.6 Rita figur!

a) Man ser att $P(B) = P(B \cap A) + P(B \cap A^{\star})$ där $P(B \cap A^{\star}) = 1 - P(B^{\star} \cup A) = 1 - 7/8 = 1/8$. Alltså, $P(B) = 1/9 + 1/8 = 17/72$.

b) Nu är

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{17}{72} - \frac{1}{9} = \frac{33}{72} = \frac{11}{24}.$$

2.7 Rita figur!

$$P(A \cup B) = P(A \cup (A^{\star} \cap B)) = P(A) + P(A^{\star} \cap B) = 0.5 + 0.1 = 0.6,$$

eftersom händelserna är disjunkta.

2.8 Rita figur! Händelsen A kan delas in i två disjunkta mängder: $A \cap B$ och $A \cap B^{\star}$. Alltså är $P(A) = P(A \cap B) + P(A \cap B^{\star})$ vilket omformas till

$$P(A \cap B^{\star}) = P(A) - P(A \cap B).$$

På samma sätt är $P(B \cap A^{\star}) = P(B) - P(A \cap B)$, och då $A \cap B^{\star}$ och $B \cap A^{\star}$ är disjunkta, så

$$P((A \cap B^{\star}) \cup (B \cap A^{\star})) = P(A \cap B^{\star}) + P(B \cap A^{\star}) = P(A) + P(B) - 2P(A \cap B).$$

2.9

a)

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

på samma sätt

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

och så vidare...

$$\leq \quad P(A_1) + P(A_2) + \cdots + P(A_n)$$

b)

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

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

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

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

och $P(A) = g/m = 1/3^{13} = 3^{-13} \approx 6.3 \cdot 10^{-7}$.

b) Låt B vara händelsen att de 12 första matcherna är rätt. Antalet element i B ges av

$$g = |B| = 1 \cdot 1 \cdots 1 \cdot 3 = 1^{12} \cdot 3 = 3$$

ty de första tolv matcherna kan endast tippas på ett sätt och sista matchen på godtyckligt sätt. Alltså $P(B) = g/m = 3/3^{13} = 3^{-12} \approx 1.9 \cdot 10^{-6}$. Notera att detta problem är ekvivalent med att få alla rätt då en stryktipsrad omfattar 12 matcher.

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

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

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 $P(A) = g/m = 26/3^{13} = 26 \cdot 3^{-13} \approx 1.6 \cdot 10^{-5}$.

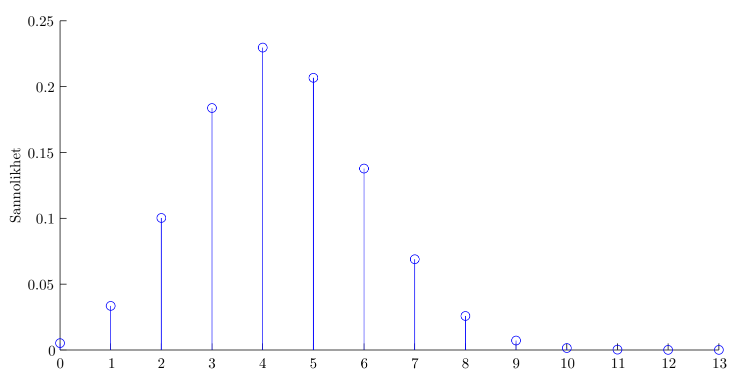
Generellt, låt D_k vara händelsen "precis k rätt". Enligt multiplikationsprincipen är antalet utfall i D_k,

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

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 $k = 0, 1, \ldots, 13$.



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

2.11 Mängden av alla stryktipsrader $\Omega = \{1, \mathrm{x}, 2\} \times \cdots \times \{1, \mathrm{x}, 2\}$ har enligt multiplikationsprincipen $m = 3 \cdot 3 \cdots 3 = 3^{12}$ element vardera med sannolikheten $1/m$.

Låt A vara händelsen att man får 0 rätt. Antalet element i A är enligt multiplikationsprincipen $g = 2 \cdot 2 \cdots 2 = 2^{13}$ eftersom varje match kan tippas fel på två sätt. Således, $P(A) = g/m = 2^{12}/3^{12} = (2/3)^{12} \approx 0.007707$. Sannolikheten för 12 rätt beräknas på samma sätt till $1/3^{12}$ så sannolikheten för inget rätt är $2^{12} = 4096$ gånger större än sannolikheten för alla rätt.

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

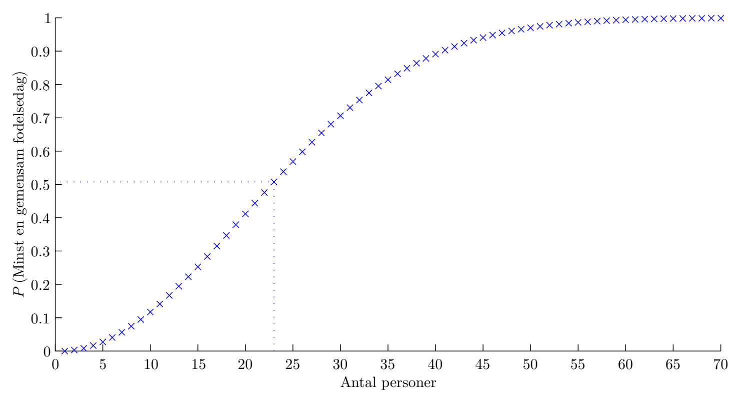
$$P(\text{Ingen gemensam}) = \frac{\text{Antalet sätt vi kan välja ut } n \text{ födelsedagar utan par}}{\text{Antalet sätt vi kan välja ut } n \text{ födelsedagar}}.$$

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

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

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%.

2.13 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

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

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

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

$$g = (\#\text{ färger}) \cdot \binom{\#\text{ följder om}}{\text{fem kort}} = 4 \cdot 9 = 36$$

så sannolikheten är $g/m = 36/\binom{52}{5}$.

c) fem kort i samma färg, är enligt multiplikationsprincipen

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

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

2.14 Slumpförsöket består av att dra 13 kort utan återläggning på måfå ur en kortlek om 52 kort. Utfallsrummet består då av de $m = \binom{52}{13} = 635013559600$ 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

$$\begin{pmatrix}\#\text{ sätt att}\\ \text{välja 5 av}\\ \text{färg 1}\end{pmatrix} \cdot \begin{pmatrix}\#\text{ sätt att}\\ \text{välja 3 av}\\ \text{färg 2}\end{pmatrix} \cdot \begin{pmatrix}\#\text{ sätt att}\\ \text{välja 3 av}\\ \text{färg 3}\end{pmatrix} \cdot \begin{pmatrix}\#\text{ sätt att}\\ \text{välja 2 av}\\ \text{färg 4}\end{pmatrix} = 8211173256$$

så det som återstår är att bestämma på hur många olika sätt man kan fördela $\heartsuit$, $\diamondsuit$, $\spadesuit$ och $\clubsuit$ 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

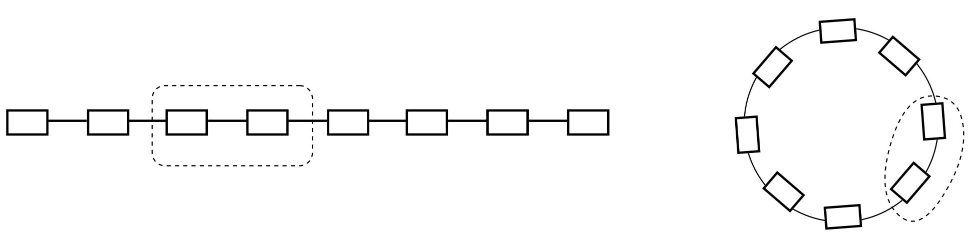
$$\binom{4}{2} \cdot 2 \cdot 1 = 12$$

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

2.15 Sannolikheten att de två maskinerna står bredvid varandra kan beräknas enligt

$$P\,(\text{Bredvid varandra}) = \frac{\text{Antalet sätt att välja ett par}}{\text{Antalet sätt att välja två}}$$

Betrakta figurerna



Antalet sätt som man kan välja ut två maskiner är $\binom{8}{2} = 28$. Antalet par i den första konfigurationen är 7 och i den andra 8, så sannolikheterna blir $7/28 = 1/4$ och $8/28 = 2/7$ respektive.

2.16 Betrakta den hög där $\diamondsuit$ ligger. Sannolikheten att $\heartsuit$ ligger i samma hög är $1/3$ (ett kort av tre möjliga), dvs. med sannolikhet $2/3$ ligger de i olika högar.

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

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

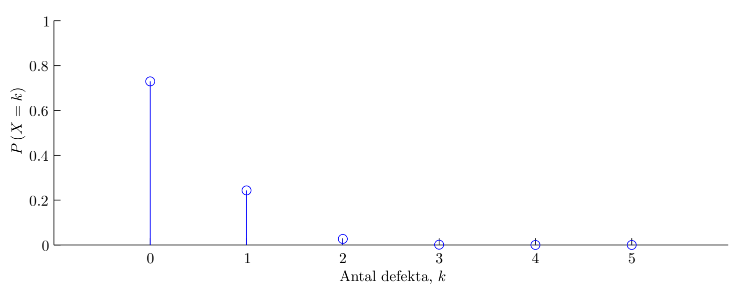
På samma sätt kan man räkna ut

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

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

$$P\left(X = k\right) = \frac{\begin{pmatrix}\text{\#sätt att välja} \\ k \text{ bland de de-} \\ \text{fekta}\end{pmatrix}\begin{pmatrix}\text{\#sätt att välja} \\ 5-k \text{ bland de} \\ \text{hela}\end{pmatrix}}{\text{\#sätt att välja 5 bland alla}} = \frac{\binom{6}{k}\binom{94}{5-k}}{\binom{100}{5}}.$$

för $k = 0, 1, \ldots, 5$



2.18 Låt urnan innehålla s svarta och v vita kulor och låt $N = s + v$. Andelen svarta är $p = s/N$ och andelen vita $v/N = 1 - p$. Dragning av n kulor (med återläggning) kan ske på N^n sätt, ty vid varje dragning har man N kulor att välja på. Av dessa är antalet dragningar av n kulor som innehåller precis k svarta

$$\begin{pmatrix}\text{Välj ut vilka } k \text{ dragningar} \\ \text{av } n \text{ som skall ge svart ku-} \\ \text{la}\end{pmatrix}\binom{\text{Välj ut } k}{\text{svarta}}\binom{\text{Välj ut}}{n-k \text{ vita}} = \binom{n}{k} \cdot s^k \cdot v^{n-k}.$$

a) Sannolikheten för exakt k svarta är

$$P\left(\text{exakt } k \text{ stycken}\right) = \frac{\binom{n}{k} \cdot s^k \cdot v^{n-k}}{N^n} = \binom{n}{k}\left(\frac{s}{N}\right)^k \cdot \left(\frac{v}{N}\right)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k},$$

för $k = 0, 1, \ldots, n$.

b) Nu är

$$\begin{aligned} P\,(\text{någon blir fel}) &= P\,(\text{minst en svart}) = 1 - P\,(\text{ingen svart}) = 1 - \binom{n}{0} p^0 (1-p)^{n-0} \\ &= 1 - (1-p)^n \end{aligned}$$

c) och

$$\begin{aligned} P\,(\text{högst en fel}) &= P\,(\text{0 eller 1 svart}) = P\,(\text{0 svarta}) + P\,(\text{1 svart}) \\ &= \binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1} = (1-p)^n + np(1-p)^{n-1}. \end{aligned}$$

d) Med $n = 3$ och $p = 0.10$ fås de numeriska svaren 0.271 och 0.972.

2.19 Att singla ett mynt $2n$ gånger kan enligt multiplikationsprincipen ge $2 \cdot 2 \cdots 2 = 2^{2n}$ olika resultat. Antalet av dessa som innehåller precis n klave (och n kronor) är att bland $2n$ mynt välja ut vilka n som är klavar. Detta kan göras på $\binom{2n}{n}$ sätt så sannolikheten för lika många kronor som klavar i en sekvens om $2n$ singlingar är

$$\begin{aligned} P\,(\text{lika många}) &= \binom{2n}{n} \Big/ 2^{2n} = \frac{(2n)!}{n!(2n-n)!} \cdot 2^{-2n} = \frac{(2n)!}{n!n!} \cdot 2^{-2n} \\ &\approx \frac{\sqrt{2\pi 2n}(2n)^{2n} \mathrm{e}^{-2n}}{(\sqrt{2\pi n} n^n \mathrm{e}^{-n})^2} \cdot 2^{-2n} = \frac{2\sqrt{\pi n} 2^{2n} n^{2n} \mathrm{e}^{-2n}}{2\pi n n^{2n} \mathrm{e}^{-2n}} \cdot 2^{-2n} = \frac{1}{\sqrt{\pi n}}. \end{aligned}$$

2.20 Låt s vara antalet svarta och v antalet vita kulor i urnan. Vi drar en kula på måfå tills vi får en svart kula.

Med återläggning. Om varje dragen kula återförs till urnan är sannolikheten att man får en svart kula $s/(s+v)$ i varje dragning. Händelsen att vi måste dra fler än k kulor är händelsen att de k första kulorna är vita. Sannolikheten för detta är

$$P\,(\text{fler än } k) = \underbrace{\left(\frac{v}{s+v}\right)\left(\frac{v}{s+v}\right)\cdots\left(\frac{v}{s+v}\right)}_{k \text{ stycken}} = \left(\frac{v}{s+v}\right)^k,$$

för $k = 0, 1, 2, \ldots$. Händelsen att vi måste dra precis k kulor är händelsen att man sett sekvensen $(k-1)$ vita kulor följt av en svart. Sannolikheten för sekvensen är

$$P\,(\text{precis } k) = \underbrace{\left(\frac{v}{s+v}\right)\left(\frac{v}{s+v}\right)\cdots\left(\frac{v}{s+v}\right)}_{k-1 \text{ stycken}}\left(\frac{s}{s+v}\right) = \left(\frac{v}{s+v}\right)^{k-1}\left(\frac{s}{s+v}\right),$$

för $k = 1, 2, 3, \ldots$.

Utan återläggning. Händelsen att vi måste dra fler än k kulor är händelsen att de k första kulorna är vita. Vi kan välja ut k kulor bland $s+v$ stycken på $\binom{s+v}{k}$ sätt. Antalet sätt av dessa som enbart innehåller vita kulor är $\binom{v}{k}$ så sannolikheten att de k första är vita är

$$P\,(\text{fler än } k) = \binom{v}{k} \Big/ \binom{s+v}{k},$$

för $k = 0, 1, \ldots, v$. Händelsen att vi måste dra precis k kulor är händelsen att man sett sekvensen $(k-1)$ vita kulor följt av en svart.

Alternativ 1: Vi kan välja ut k kulor bland $s+v$ stycken på $\binom{s+v}{k}$ sätt. Antalet sätt som vi kan välja ut k kulor så att $k-1$ är vita och 1 är svart är $\binom{v}{k-1}\binom{s}{1}$. Detta är oavsett ordning så av dessa sätt är det bara andelen $1/k$ stycken som har den svarta sist. Alltså är

$$P\,(\text{exakt } k) = \frac{\binom{v}{k-1}\binom{s}{1}\frac{1}{k}}{\binom{s+v}{k}} = \frac{\binom{v}{k-1}}{\binom{s+v}{k}}\,\frac{s}{k}.$$

Alternativ 2: Välj kula en i taget. Första kulan har sannolikhet $v/(s+v)$ att vara vit. Om första är vit så är nästa vit med sannolikhet $(v-1)/(s+v-1)$ eftersom det är $s+v-1$ kulor kvar i urnan varav $v-1$ är vita. Skall man välja precis $k-1$ vita följt av en svart får man

$$\begin{aligned}
P\,(\text{exakt } k) &= \frac{v}{s+v}\,\frac{v-1}{s+v-1}\cdots\frac{v-(k-1)+1}{s+v-(k-1)+1}\cdot\frac{s}{s+v-(k-1)}\\
&= s\frac{v!}{(v-(k-1))!}\cdot\frac{(s+v-k)!}{(s+v)!}\\
&= s\frac{v!}{(v-(k-1))!(k-1)!}(k-1)!\cdot\frac{(s+v-k)!k!}{(s+v)!}\,\frac{1}{k!}\\
&= s\binom{v}{k-1}(k-1)!\left(\binom{s+v}{k}k!\right)^{-1} = \frac{s}{k}\,\frac{\binom{v}{k-1}}{\binom{s+v}{k}}.
\end{aligned}$$

Alternativ 3: Att välja ut bara vita kulor för de $k-1$ första dragningarna har sannolikhet

$$\binom{v}{k-1}\Big/\binom{s+v}{k-1}$$

för $k=1,2,\ldots,v+1$. När man valt ut dessa $k-1$ vita finns det $s+v-(k-1)$ kulor kvar i urnan varav s är svarta så sannolikheten att de $k-1$ vita följs av en svart är $s/(s+v-(k-1))$. Slutligen

$$P\,(\text{precis } k) = \frac{\binom{v}{k-1}}{\binom{s+v}{k-1}}\,\frac{s}{s+v-(k-1)} = \frac{s}{k}\,\frac{\binom{v}{k-1}}{\binom{s+v}{k}},$$

för $k=1,\ldots,v+1$.

2.21

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

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

2.22 Låt A_k vara händelsen ”ingen sexa bland de k första kasten” för $k=1,2,\ldots$ Då är

$$\begin{aligned}
P\,(A_k) &= P\,(\{\text{ej 6:a i kast 1}\}\cap\{\text{ej 6:a i kast 2}\}\cap\cdots\cap\{\text{ej 6:a i kast } k\}) = \{\text{ober}\}\\
&= P\,(\{\text{ej 6:a}\})\cdots P\,(\{\text{ej 6:a}\}) = \frac{5}{6}\cdots\frac{5}{6} = \left(\frac{5}{6}\right)^k
\end{aligned}$$

för $k = 1, 2, \ldots$ Vidare så är enligt definitionen av betingad sannolikhet

$$P\left(A_{k+n}|A_n\right) = \frac{P\left(A_{k+n} \cap A_n\right)}{P\left(A_n\right)}$$

Om A_{k+n} inträffar så inträffar även A_n, dvs $A_{k+n} \cap A_n = A_{k+n}$

$$P\left(A_{k+n}|A_n\right) = \frac{P\left(A_{k+n}\right)}{P\left(A_n\right)} = \frac{(5/6)^{k+n}}{(5/6)^n} = (5/6)^k = P\left(A_k\right).$$

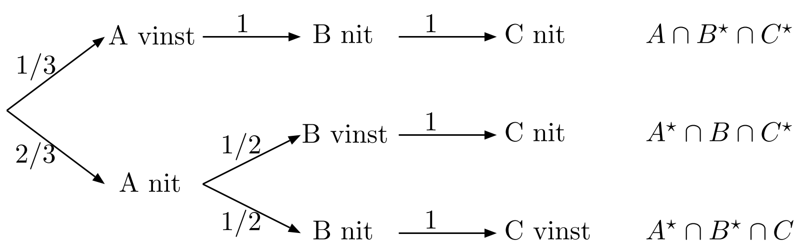
2.23 Låt A, B och C vara händelsen att person A, B och C får vinstlotten. Notera att händelserna är disjunkta och $P\left(A \cup B \cup C\right) = 1$. Eftersom A tar första lotten är $P\left(A\right) = 1/3 = P\left(A \cap B^\star \cap C^\star\right)$.

$$P\left(B\right) = P\left(B|A\right)P\left(A\right) + P\left(B|A^\star\right)P\left(A^\star\right) = 0\,\frac{1}{3} + \frac{1}{2}\,\frac{2}{3} = \frac{1}{3}$$

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

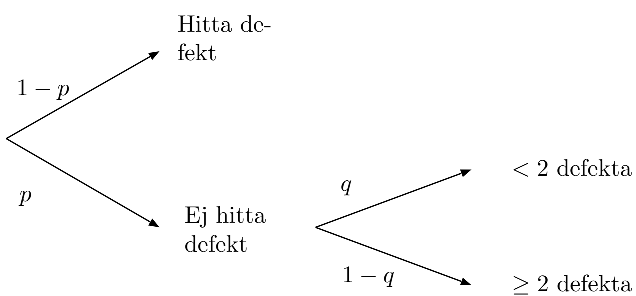
$$P\left(C\right) = P\left(C|B\right)P\left(B\right) + P\left(C|B^\star\right)P\left(B^\star\right) = 0\,\frac{1}{3} + \frac{1}{2}\,\frac{2}{3} = \frac{1}{3}.$$

Resonemanget illustreras enkelt med ett träddiagram:



Träddiagram som visar de tre möjliga fallen: A vinner, B vinner och C vinner.

2.24 Betrakta träddiagrammet över testförfarandet.



Träddiagram över testförfarandet.

Vid första steget undersöks 5 på måfå utvalda enheter av 50. Vi kan välja ut 5 av 50 på $\binom{50}{5} = 2118760$ sätt. Av dessa är antalet sätt som man enbart väljer felfria enheter

$$\begin{pmatrix}\text{\# sätt att}\\ \text{välja 0 bland}\\ \text{defekta}\end{pmatrix}\begin{pmatrix}\text{\# sätt att}\\ \text{välja 5 bland}\\ \text{hela}\end{pmatrix} = \binom{5}{0}\binom{45}{5} = 1221759$$

så sannolikheten att inte finna någon defekt vid första steget är

$$p = P\left(\text{ingen defekt}\right) = \frac{\binom{5}{0}\binom{45}{5}}{\binom{50}{5}} = 0.5766.$$

Om man inte funnit någon defekt inleds steg två där 10 av de kvarvarande 45 enheterna undersöks. Dessa 10 kan väljas ut på $\binom{45}{10} = 3190187286$ sätt. Antalet sätt av dessa som motsvarar ett val av 0 eller 1 defekt är

$$\binom{5}{0}\binom{40}{10} + \binom{5}{1}\binom{40}{9} = 847660528 + 1367194400 = 2214854928$$

så sannolikheten q för färre än två defekta är

$$q = \frac{\binom{5}{0}\binom{40}{10} + \binom{5}{1}\binom{40}{9}}{\binom{45}{10}} = \frac{2214854928}{3190187286} = 0.6943.$$

Sannolikheten för att paritet accepteras är $p \cdot q = 0.40034$ och partiet avvisas med sannolikhet 0.60.

2.25 Från en urna med v vita och $s = N - v$ svarta kulor drar man på måfå utan återläggning tills man får en svart kula. Bestäm med successiv betingning sannolikheten att man får den första svarta kulan i dragning k, $k = 1, 2, \ldots, v + 1$. Då har man dragit sekvensen $k - 1$ vita kulor följt av en svart.

Alternativ 1: Betingning på varje dragning. Välj kula en i taget. Första kulan har sannolikhet $v/(s+v)$ att vara vit. Givet att den första är vit så är den betingade sannolikheten att nästa är vit $(v-1)/(s+v-1)$ eftersom det är $s + v - 1$ kulor kvar i urnan varav $v - 1$ är vita. Har man dragit k vita kulor är även kula $k + 1$ vit med sannolikhet $(v - k)/(s + v - k)$. Skall man välja precis $k - 1$ vita följt av en svart får man

$$\begin{aligned}
P\,(\text{exakt } k) &= \frac{v}{s+v}\,\frac{v-1}{s+v-1}\cdots\frac{v-(k-1)+1}{s+v-(k-1)+1}\cdot\frac{s}{s+v-(k-1)} \\
&= s\frac{v!}{(v-(k-1))!}\cdot\frac{(s+v-k)!}{(s+v)!} \\
&= s\frac{v!}{(v-(k-1))!(k-1)!}(k-1)!\cdot\frac{(s+v-k)!k!}{(s+v)!}\,\frac{1}{k!} \\
&= s\binom{v}{k-1}(k-1)!\left(\binom{s+v}{k}k!\right)^{-1} = \frac{s}{k}\,\frac{\binom{v}{k-1}}{\binom{s+v}{k}},
\end{aligned}$$

för $k = 1, \ldots, v + 1$.

Alternativ 2: Betinga på dragningen av de vita kulorna. Att välja ut bara vita kulor för de $k-1$ första dragningarna har sannolikhet

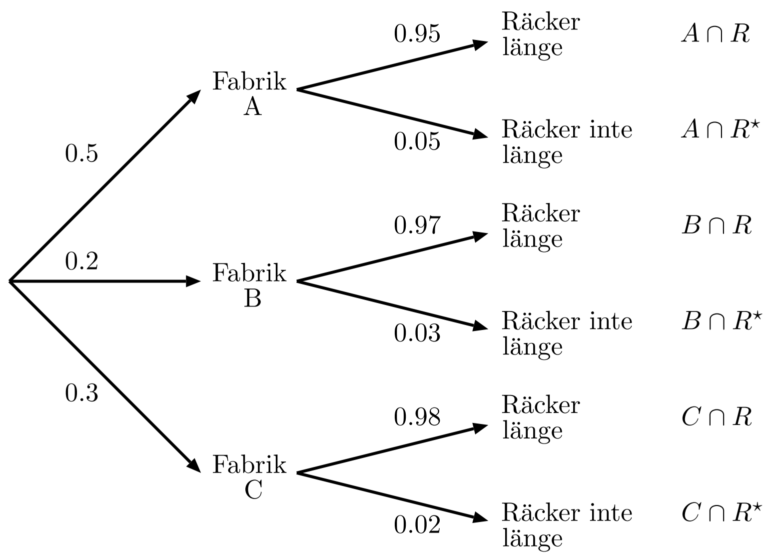
$$\binom{v}{k-1}\Big/\binom{s+v}{k-1}$$

för $k = 1, 2, \ldots, v + 1$. När man valt ut dessa $k - 1$ vita finns det $s + v - (k - 1)$ kulor kvar i urnan varav s är svarta så sannolikheten att de $k - 1$ vita följs av en svart är $s/(s + v - (k - 1))$. Slutligen

$$P\,(\text{precis } k) = \frac{\binom{v}{k-1}}{\binom{s+v}{k-1}}\,\frac{s}{s+v-(k-1)} = \frac{s}{k}\,\frac{\binom{v}{k-1}}{\binom{s+v}{k}},$$

för $k = 1, \ldots, v + 1$.

2.26 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. Enklast illustreras sambandet mellan händelserna och deras betingade sannolikheter i ett träddiagram 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.



Alltså,

$$\begin{aligned} P(A \cap R) &= P(R|A)\,P(A) = 0.95 \cdot 0.5 = 0.475 \\ P(B \cap R) &= P(R|B)\,P(B) = 0.97 \cdot 0.2 = 0.194 \\ P(C \cap R) &= P(R|C)\,P(C) = 0.98 \cdot 0.3 = 0.294 \end{aligned}$$

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 så är

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

och

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

En jämförelse mellan $P(A|R)$, $P(A|R^\star)$ och $P(A)$ ger vid handen att $P(A|R)$ skiljer sig "mindre" från $P(A) = 0.5$ än vad $P(A|R^\star)$ gör.

Notera att

$$P(A) = P(A|R)\,P(R) + P(A|R^\star)\,P(R^\star)$$

och att alla sannolikheter är på intervallet $[0,\ 1]$. Man kan se $P(A)$ som ett viktat medelvärde av $P(A|R)$ och $P(A|R^\star)$ med vikterna $P(R)$ och $P(R^\star)$ respektive.

Detta medför att om $P(A|R)$ är mindre än $P(A)$ så måste detta balanseras av att $P(A|R^\star)$ är större än $P(A)$ och vice versa. Om $P(R)$ är nära ett, så att $P(A|R)$ har en stor vikt, så måste en liten avvikelse mellan $P(A|R)$ och $P(A)$ balanseras av en större (motsatt) avvikelse mellan $P(A|R^\star)$ och $P(A)$.

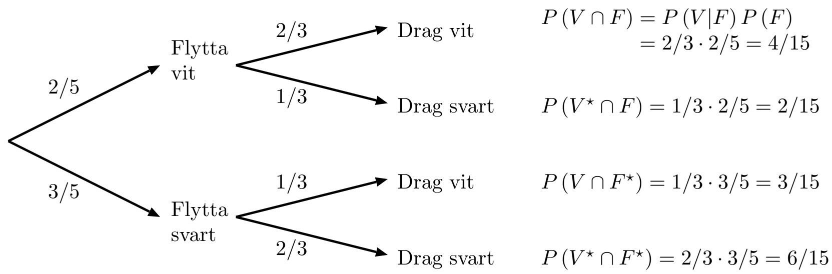
2.27 Låt F vara händelsen "den flyttade kulan är vit" och V händelsen "den dragna kulan är vit".

a) Då är $F^\star$ händelsen "flyttad kula svart" och

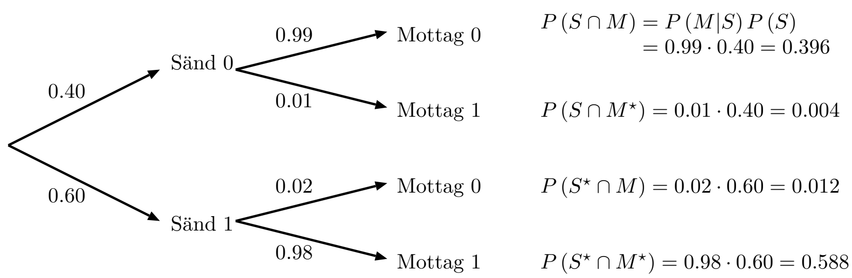
$$\begin{aligned} P(V) &= P((V \cap F) \cup (V \cap F^\star)) = P(V \cap F) + P(V \cap F^\star) \\ &= P(V|F)\,P(F) + P(V|F^\star)\,P(F^\star) = \frac{2}{3} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{3}{5} = \frac{7}{15}. \end{aligned}$$

b) Vi söker nu $P(F^\star|V)$.

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



2.28 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 träddiagrammet nedan.



De sökta sannolikheterna är (kolla figuren!)

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

och

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

2.29 Låt p vara person A:s träffsannolikhet och q vara person B:s träffsannolikhet. Personerna träffar oberoende av varandra och A börjar skjuta. Sannolikheten att A träffar först är ett jämnt antal missar följt av en träff.

$$\begin{aligned}
P(\text{A träff}) &+ P(\text{A miss, B miss, A träff}) + P(\text{A miss, B miss, A miss, B miss, A träff}) + \cdots \\
&= p + (1-p)(1-q)p + (1-p)(1-q)(1-p)(1-q)p + \cdots \\
&= \sum_{k=0}^{\infty} p(\underbrace{(1-p)(1-q)}_{=s})^k = p\sum_{k=0}^{\infty} s^k = \{\text{geometrisk serie}\} \\
&= p\,\frac{1}{1-s} = p\,\frac{1}{1-(1-p)(1-q)}.
\end{aligned}$$

Med $p = 1/10$ och $q = 1/9$ blir sannolikheten

$$\frac{p}{1-(1-p)(1-q)} = \frac{1/10}{1-(1-1/10)(1-1/9)} = \frac{1}{10-(10-1)(1-1/9)} = \frac{1}{2}.$$

2.30 Låt p_k vara sannolikheten att A vinner givet att A har k kronor. Betingning på hur första singlingen utfaller ger

$$\begin{aligned} p_k &= P\left(\text{A vinner med } k+1 \text{ kronor}|\text{A vinner första}\right) P\left(\text{A vinner första}\right) \\ &+ P\left(\text{A vinner med } k-1 \text{ kronor}|\text{A förlorar första}\right) P\left(\text{A förlorar första}\right) \\ &= p_{k+1}\frac{1}{2} + p_{k-1}\frac{1}{2}. \end{aligned}$$

Vi skall lösa $p_k = \frac{1}{2}(p_{k-1} + p_{k+1})$, $k = 1, 2, \ldots, a+b-1$ eller

$$p_{k+2} - 2p_{k+1} + p_k = 0$$

för $k = 0, 1, \ldots, a+b-2$, under bivillkor att $p_0 = 0$ (om A inte har några pengar förlorar A) och $p_{a+b} = 1$ (A vinner om B inte har några pengar). Den homogena differensekvationen har allmänna lösningar på formen $p_k = Ac^k$ för konstanter A och c. Insatt i differensekvationen fås

$$Ac^{k+2} - 2Ac^{k+1} + Ac^k = Ac^k(c^2 - 2c + 1) = 0.$$

Den karaktäristiska ekvationen $c^2 - 2c + 1 = (c-1)^2 = 0$ har dubbelroten $c = 1$ som lösning. Det allmänna lösningen för p_k blir då

$$p_k = (A_1 k + A_2)c^k = (A_1 k + A_2)1^k = A_1 k + A_2$$

för konstanter A_1 och A_2. Bivillkoren bestämmer konstanterna

$$\begin{aligned} 0 = p_0 = A_1 \cdot 0 + A_2 \quad &\Rightarrow \quad A_2 = 0 \\ 1 = p_{a+b} = A_1(a+b) + A_2 = A_1(a+b) + 0 \quad &\Rightarrow \quad A_1 = 1/(a+b) \end{aligned}$$

Alltså:

$$p_k = A_1 k + A_2 = \frac{1}{a+b}k.$$

för $k = 0, 1, \ldots, a+b$, och speciellt $p_a = a/(a+b)$ och $1 - p_a = 1 - a/(a+b) = b/(a+b)$.

2.31

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. $P(T|S) = 10/11$.

2.32

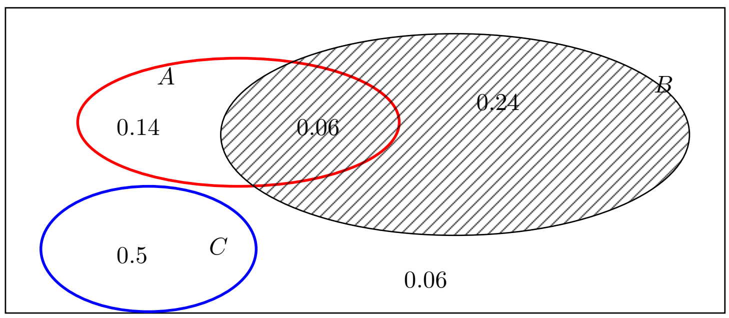
a) $(A \cup B)$ och C är disjunkta, dvs $(A \cup B) \cap C = \emptyset$ så $P((A \cup B) \cap C) = 0$.

b)

$$\begin{aligned} P(A \cup B \cup C) &= \{\text{disjunkta}\} = P(A \cup B) + P(C) = P(A) + P(B) - P(A \cap B) + P(C) \\ &= \{A \text{ och } B \text{ oberoende}\} = P(A) + P(B) - P(A)P(B) + P(C) \\ &= 0.2 + 0.3 - 0.2 \cdot 0.3 + 0.5 = 0.94. \end{aligned}$$

c)

$$\begin{aligned} P((A \cap B) \cup C) &= P(A \cap B) + P(C) - P((A \cap B) \cap C) = P(A)P(B) + P(C) \\ &= 0.2 \cdot 0.3 + 0.5 = 0.56. \end{aligned}$$



Venndiagram där C är disjunkt med de oberoende händelserna A och B

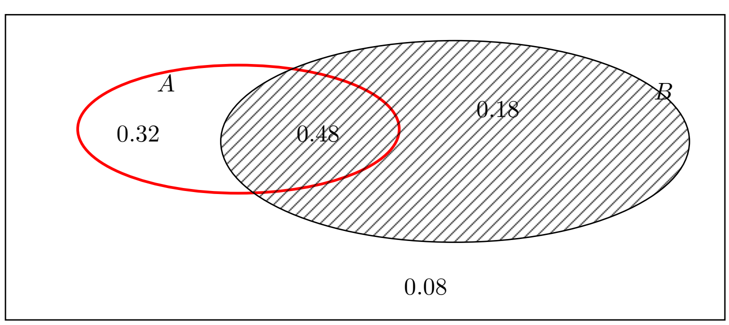
2.33 För att undersöka om A och B oberoende skall man svara på frågan om $P(A \cap B)$ är lika med $P(A)P(B)$. Nu är

$$\begin{aligned} P(A^\star \cap B) &= 1 - P(A \cup B^\star) = 1 - 0.88 = 0.12. \\ P(A \cap B^\star) &= 1 - P(A^\star \cup B) = 1 - 0.68 = 0.32. \end{aligned}$$

Likheten $P(A \cup B) = P(A^\star \cap B) + P(A \cap B) + P(A \cap B^\star)$ ger

$$\begin{aligned} P(A \cap B) &= P(A \cup B) - P(A^\star \cap B) - P(A \cap B^\star) = 0.92 - 0.12 - 0.32 = 0.48 \\ P(A) &= P(A \cap B) + P(A \cap B^\star) = 0.48 + 0.32 = 0.80 \\ P(B) &= P(A \cap B) + P(A^\star \cap B) = 0.48 + 0.12 = 0.60 \end{aligned}$$

Alltså är $P(A \cap B) = 0.48 = 0.80 \cdot 0.60 = P(A)P(B)$ och A och B är oberoende.



Venndiagram där händelserna A och B är oberoende.

2.34 Sannolikheten att en tillverkad komponent har minst ett av felen

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

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

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

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

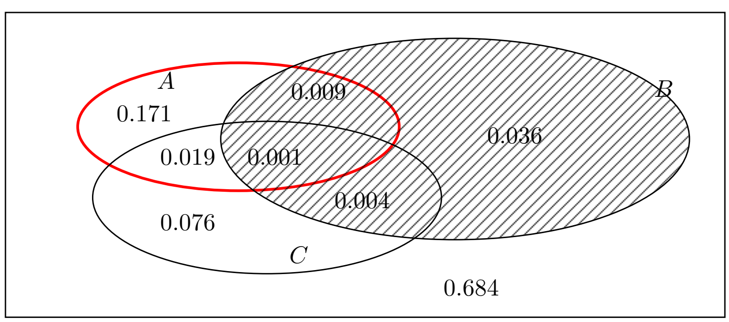
Alternativ 3: Låt X beskriva antalet fel hos produkten. $\{X = 0\} = A^\star \cap B^\star \cap C^\star$ så utnyttjandes oberoendet

$$P(X = 0) = P(A^\star)\,P(B^\star)\,P(C^\star) = 0.684$$

och på samma sätt fås

$$\begin{aligned} P(X = 1) &= P(A^\star)\,P(B^\star)\,P(C) + P(A^\star)\,P(B)\,P(C^\star) + P(A)\,P(B^\star)\,P(C^\star) = 0.283 \\ P(X = 2) &= P(A)\,P(B)\,P(C^\star) + P(A)\,P(B^\star)\,P(C) + P(A^\star)\,P(B)\,P(C) = 0.032 \\ P(X = 3) &= P(A)\,P(B)\,P(C) = 0.001 \end{aligned}$$

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



Venndiagram där händelserna motsvarande felen A, B och C är oberoende.

2.35 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 <u>inte</u> har att $P(A \cap B) = P(A)\,P(B)$. Händelserna A och B är ej oberoende.

2.36 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.

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

a) Då gäller för ett *seriesystem*

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

b) För ett *parallellsystem* har vi att

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

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

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

2.38 Ett elektronrör håller mer än t timmar med sannolikhet $\mathrm{e}^{-0.002t}$ för $t \geq 0$. Sannolikheten att ett elektronrör går sönder inom 500 timmar

$$p_{500} = 1 - \mathrm{e}^{-0.002 \cdot 500} = 1 - \mathrm{e}^{-1} = 0.6321.$$

Sannolikheten att två oberoende elektronrör går sönder inom 500 timmar blir således $p_{500} \cdot p_{500} = (1 - \mathrm{e}^{-1})^2 = 0.3996$.

Att ett elektronrör håller mer än 1000 timmar ges av $q = \mathrm{e}^{-0.002 \cdot 1000} = \mathrm{e}^{-2} = 0.1353$. Den sökta sannolikheten är

$$\begin{aligned} &P(\text{(Rör 1<500 och Rör2>1000) eller (Rör 1>1000 och Rör2<500)}) \\ &\quad = P(\text{Rör 1<500 och Rör2>1000}) + P(\text{Rör 1>1000 och Rör2<500}) \\ &\quad = P(\text{Rör 1<500})\, P(\text{Rör2>1000}) + P(\text{Rör 1>1000})\, P(\text{Rör2<500}) \\ &\quad = pq + qp = 2pq = 0.1711. \end{aligned}$$

2.39 (Se även uppgift 2.37.) 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 $\{\text{parallellsystem fungerar}\} = A \cup B = (A^\star \cap B^\star)^\star$.

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

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

Motsvarande för parallellsystem 2 är

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

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

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

2.40 Låt A och B vara händelserna att resp. vattenkraftverk fungerar och C att värmekraftverket fungerar. Vi vet att $P(A) = 0.98$, $P(B) = 0.98$, $P(C) = 0.9$ samt att händelserna är oberoende.

Följande 8 scenarion är möjliga. Ett 'x' markerar att händelsen har inträffat. Sannolikheten i rad 3, säg, bestäms som $P(A^\star \cap B \cap C^\star) = \{\text{ober.}\} = P(A^\star)\, P(B)\, P(C^\star) = (1 - 0.98) \cdot 0.98 \cdot (1 - 0.90) = 0.00196$.

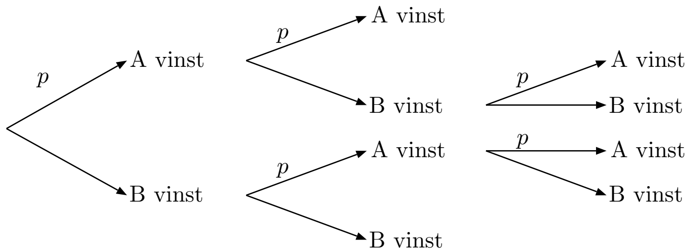
A	B	C	sannolikhet	effekt
			0.00004	$0+0+0=0$
		x	0.00036	$0+0+20=20$
	x		0.00196	$0+10+0=10$
	x	x	0.01764	$0+10+20=30$
x			0.00196	$10+0+0=10$
x		x	0.01764	$10+0+20=30$
x	x		0.09604	$10+10+0=20$
x	x	x	0.86436	$10+10+20=40$

Modellerar vi de möjliga effekterna som utfall tillskriver vi effekterna (utfallen) $\{0,\ 10,\ 20,\ 30,\ 40\}$ sannolikheter:

utfall	sannolikhet
0	0.00004
10	$0.00196 + 0.00196 = 0.00392$
20	$0.00036 + 0.09604 = 0.09640$
30	$0.01764 + 0.01764 = 0.03528$
40	0.86436

Den maximala effekten är 40MW. Händelsen att minst hälften av den maximala effekten är tillgänglig består av utfallen $\{20,\ 30,\ 40\}$ med sannolikhet $0.09640 + 0.03528 + 0.86436 = 0.99604$.

2.41 Rita ett träddiagram. Låt $p = 0.6$ vara sannolikheten att A vinner en match, dvs $1 - p = 0.40$ är sannolikheten att B vinner.



Träddiagram för de möjliga utgångarna när A och B spelar tennis

Matchen är slutar efter 2 set om A vinner båda eller B vinner båda. Detta sker med sannolikhet

$$p \cdot p + (1-p) \cdot (1-p) = 0.6^2 + 0.4^2 = 0.52.$$

Således slutar matchen efter tre set med sannolikhet $1 - 0.52 = 0.48$. Vidare A vinner med sannolikhet (se träddiagrammet)

$$p\,p + p(1-p)p + (1-p)p\,p = (3-2p)p^2 = 0.648.$$

2.42 Om det sjätte skottet träffar måltavla tre har man bland de fem första skotten haft två träffar och det sjätte skottet är en träff. Alltså, sannolikheten är

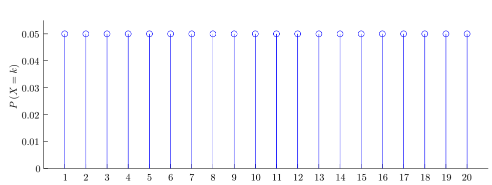
$$\begin{aligned} P(\{\text{2 träffar och 3 missar bland de 5 första}\} \cap \{\text{sista träff}\}) &= \{\text{oberoende skott}\} \\ &= \left(\binom{5}{2}(0.40)^2(1-0.40)^3\right)0.40 \\ &= 0.1382. \end{aligned}$$

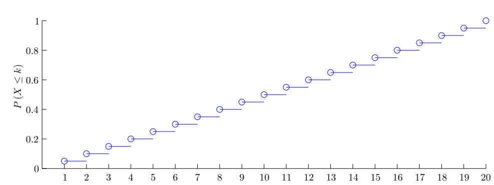
3.1 Låt X beteckna det nummer där lyckohjulet med n nummer stannar. De möjliga värdena på X är $S_X = \{1, 2, \ldots, n\}$. Ett rättvist hjul har samma sannolikhet för alla nummer, dvs $p_X(k) = P(X = k) = p$ beror ej av k. Eftersom

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

så är $p = 1/n$, dvs $P(X = k) = 1/n$ för $k \in S_X$. 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 Låt den stokastiska variabeln X ha sannolikheter

$$P(X = k) = \frac{m^k}{k!}\mathrm{e}^{-m}$$

för något tal $m > 0$ och $k = 0, 1, 2, \ldots$. Vi söker sannolikheten $P(2 < X < 5)$. Alltså,

$$\begin{aligned} P(2 < X < 5) &= P(\{X = 3\} \cup \{X = 4\}) = P(X = 3) + P(X = 4) \\ &= \frac{m^3}{3!}\mathrm{e}^{-m} + \frac{m^4}{4!}\mathrm{e}^{-m} = \frac{4^3}{6}\mathrm{e}^{-4} + \frac{4^4}{24}\mathrm{e}^{-4} = \frac{128}{6}\mathrm{e}^{-4}. \end{aligned}$$

3.3 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 $S_X = \{0, 5, 20, 100\}$. 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.

3.4 De möjliga värdena på X är $S_X = \{1, 2, 3, \ldots\}$, ett uppräkneligt oändligt antal värden. Händelsen $X = k$, där $k = 1, 2, \ldots$, är händelsen att man sett sekvensen $(k-1)$ defekta enheter följt av 1 hel enhet. Sannolikheten för en sådan sekvens är, med hjälp av oberoendet

$$P(X = k) = P(\{\text{(k-1) defekta följt av 1 hel}\}) = \underbrace{(1-p)\cdot(1-p)\cdots(1-p)}_{(k-1)\text{ st}} \cdot p = (1-p)^{k-1}p$$

för $k = 1, 2, 3, \ldots$

3.5 Låt A_k stå för händelsen att terminal k används, $k = 1, 2, 3$. Den stokastiska variabeln X kan anta värdena 0, 1, 2 och 3. Vi erhåller att

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

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 $\{A_1 \cap A_2 \cap A_3^\star\} \cup \{A_1 \cap A_2^\star \cap A_3\} \cup \{A_1^\star \cap A_2 \cap A_3\}$. Vi erhåller

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

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

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

3.6 Händelsen A består av utfallen {⚀, ⚂, ⚄}, händelsen B av {⚀, ⚁, ⚂} samt C av {⚀, ⚁}. Varje tärningsutfall har sannolikhet 1/6 och följande schema ger vinsten som funktion av utfallet:

utfall	A	B	C	vinst	netto
⚀	x	x	x	$3+2+1=6$	3
⚁		x	x	$2+1=3$	0
⚂	x	x		$3+2=5$	2
⚃				0	-3
⚄	x			3	0
⚅				0	-3

Låt X beskriva spelarens nettovinst. De möjliga värdena på X är $\{-3,\ 0,\ 2,\ 3\}$. Sannolikhetsfunktionen $p_X(x)$ ges av

$$\begin{aligned} p_X(-3) &= P(X=0) = 2/6 \\ p_X(0) &= P(X=3) = 2/6 \\ p_X(2) &= P(X=5) = 1/6 \\ p_X(3) &= P(X=6) = 1/6 \end{aligned}$$

och $p_X(x) = 0$ för övrigt.

3.7 Låt A och B vara händelserna att resp. vattenkraftverk fungerar och C att värmekraftverket fungerar. Vi vet att $P(A) = 0.98$, $P(B) = 0.98$, $P(C) = 0.9$ samt att händelserna är oberoende.

Följande 8 scenarion är möjliga. Ett 'x' markerar att händelsen har inträffat. Sannolikheten i rad 3, säg, bestäms som $P(A^\star \cap B \cap C^\star) = \{\text{ober.}\} = P(A^\star)\,P(B)\,P(C^\star) = (1-0.98)\cdot 0.98\cdot(1-0.90) = 0.00196$.

A	B	C	sannolikhet	antal fungerande	effekt
			0.00004	0	$0+0+0=0$
		x	0.00036	1	$0+0+20=20$
	x		0.00196	1	$0+10+0=10$
	x	x	0.01764	2	$0+10+20=30$
x			0.00196	1	$10+0+0=10$
x		x	0.01764	2	$10+0+20=30$
x	x		0.09604	2	$10+10+0=20$
x	x	x	0.86436	3	$10+10+20=40$

Låt X beskriva tillgänglig effekt och Y antalet fungerande kraftverk.

De möjliga värdena på X är $\{0,\ 10,\ 20,\ 30,\ 40\}$ med sannolikhetsfunktionen:

utfall, x	$p_X(x)$
0	0.00004
10	$0.00196+0.00196=0.00392$
20	$0.00036+0.09604=0.09640$
30	$0.01764+0.01764=0.03528$
40	0.86436

och $p_X(x) = 0$ för övrigt.

De möjliga värdena på Y är $S_Y = \{0,1,2,3\}$ med sannolikhetsfunktionen:

utfall, x	$p_Y(x)$
0	0.00004
1	$0.00036+0.00196+0.00196=0.00428$
2	$0.01764+0.01764+0.09604=0.13132$
3	0.86436

och $p_Y(x) = 0$ för övrigt.

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

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

så

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

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

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

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

3.9 a) Låt X vara ffg(p)-fördelad. Då är $P(X = k) = (1-p)^{k-1} p$ för $k = 1, 2, 3, \ldots,$. Vidare

$$\begin{aligned} P(X \text{ udda}) &= \sum_{\text{udda } k} P(X = k) = \sum_{i=0}^{\infty} P(X = 2i + 1) = \sum_{i=0}^{\infty} (1-p)^{2i+1-1} p \\ &= p \sum_{i=0}^{\infty} (\underbrace{(1-p)^2}_{=q})^i = p \sum_{i=0}^{\infty} q^i = p \frac{1}{1-q} = p \frac{1}{1-(1-p)^2} = \frac{1}{2-p}. \end{aligned}$$

Notera att $p = 1$ ger $P(X \text{ udda}) = 1$ och $p \to 0$ ger $P(X \text{ udda}) \to 1/2$.

b) Låt X vara Po(m)-fördelad. Då är $P(X = k) = \frac{m^k}{k!} \mathrm{e}^{-m}$ för $k = 0, 1, 2, \ldots,$. Vidare

$$P(X \text{ udda}) = \sum_{\text{udda } k} P(X = k) = \sum_{\text{udda } k} \frac{m^k}{k!} \mathrm{e}^{-m} = \mathrm{e}^{-m} \sum_{\text{udda } k} \frac{m^k}{k!}.$$

Vi vet att

$$\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

för alla x. Alltså är

$$\mathrm{e}^x + \mathrm{e}^{-x} = \sum_{k=0}^{\infty} \frac{x^k}{k!} + \sum_{k=0}^{\infty} \frac{(-x)^k}{k!} = \sum_{k=0}^{\infty} \frac{x^k(1^k + (-1)^k)}{k!} = 2 \sum_{\text{jämna } k} \frac{x^k}{k!}$$

och

$$\mathrm{e}^x - \mathrm{e}^{-x} = \sum_{k=0}^{\infty} \frac{x^k}{k!} - \sum_{k=0}^{\infty} \frac{(-x)^k}{k!} = \sum_{k=0}^{\infty} \frac{x^k(1^k - (-1)^k)}{k!} = 2 \sum_{\text{udda } k} \frac{x^k}{k!}.$$

Utnyttjas detta får vi att

$$P(X \text{ udda}) = \mathrm{e}^{-m} \sum_{\text{udda } k} \frac{m^k}{k!} = \mathrm{e}^{-m} \frac{1}{2} \left(\mathrm{e}^m - \mathrm{e}^{-m}\right) = \frac{1}{2} \left(1 - \mathrm{e}^{-2m}\right).$$

Sannolikheten $P(X \text{ jämn}) = 1 - P(X \text{ udda})$.

3.10 De möjliga värdena på X ges av intervallet $S_X = [0, 10)$. Intervallet innehåller ett överuppräkneligt 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 sannolikheten att få vänta högst t minuter är sannolikheten att komma till stationen under ett tidsintervall om t minuter före nästa tåg av 10 möjliga minuter.

Med

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

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

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

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

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

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

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

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

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

$$f_X(t) = \frac{d}{dt}F_X(t) = 0 - \mathrm{e}^{-t/m}\frac{-1}{m} = \frac{1}{m}\mathrm{e}^{-t/m},$$

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

Alternativ 1: Med $m = 1.5$ fås

$$\begin{aligned} P(1 < X \leq 10) &= \int_1^{10} f_X(x)\,dx = \int_1^{10} \frac{1}{m}\mathrm{e}^{-x/m}dx = \left[\frac{1}{m}\mathrm{e}^{-x/m}(-m)\right]_1^{10} \\ &= \mathrm{e}^{-1/m} - \mathrm{e}^{-10/m} = 0.512. \end{aligned}$$

Alternativ 2: Med $m = 1.5$ fås

$$\begin{aligned} P(1 < X \leq 10) &= F_X(10) - F_X(1) = (1 - \mathrm{e}^{-10/m}) - (1 - \mathrm{e}^{-1/m}) \\ &= \mathrm{e}^{-1/m} - \mathrm{e}^{-10/m} = 0.512. \end{aligned}$$

3.12 Vi söker sannolikheten $P(X > 6.3)$ där X har täthetsfunktionen

$$f_X(x) = \frac{1}{m}\mathrm{e}^{-x/m}, \quad \text{då } x \geq 0.$$

för $m = 2.35$.

Alternativ 1: Direkt från täthetsfunktionen.

$$P(X > 6.3) = \int_{6.3}^{\infty} f_X(x)\,dx = \int_{6.3}^{\infty} \frac{1}{2.35}\mathrm{e}^{-x/2.35}\,dx = [-\mathrm{e}^{-x/2.35}]_{6.3}^{\infty} = \mathrm{e}^{-6.3/2.35} \approx 0.0685.$$

Alternativ 2: Via fördelningsfunktionen. För $t \geq 0$ är

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_0^t \frac{1}{m} \mathrm{e}^{-x/m} dx = 1 - \mathrm{e}^{-t/m}.$$

Nu är

$$P(X > 6.3) = 1 - P(X \leq 6.3) = 1 - F_X(6.3) = \mathrm{e}^{-6.3/m} = \mathrm{e}^{-6.3/2.35} \approx 0.0685.$$

3.13 Den stokastiska variabeln X har fördelningsfunktion

$$F_X(x) = P(X \leq x) = 1 - \mathrm{e}^{-x^2/2}$$

för $x \geq 0$. Vi söker medianen μ, den punkt sådan att

$$P(X \leq \mu) = P(X \geq \mu).$$

Men då är

$$2P(X \leq \mu) = P(X \leq \mu) + P(X \geq \mu) = 1 + \underbrace{P(X = \mu)}_{=0} = 1$$

så μ bestäms ur

$$P(X \leq \mu) = F_X(\mu) = \frac{1}{2}.$$

Med uttrycket för fördelningsfunktionen får man

$$\begin{aligned}
1/2 &= 1 - \mathrm{e}^{-\mu^2/2} \\
\mathrm{e}^{-\mu^2/2} &= 1/2 \\
-\mu^2/2 &= \ln(1/2) \\
\mu^2/2 &= \ln(2) \\
\mu &= \sqrt{2\ln(2)}.
\end{aligned}$$

Tätheten för X fås ur

$$f_X(x) = \frac{d}{dx} F_X(x) = 0 - \mathrm{e}^{-x^2/2} \frac{-1}{2} 2x = x\mathrm{e}^{-x^2/2}, \quad x \geq 0.$$

3.14 Den stokastiska variabeln X har fördelningsfunktion

$$F_X(x) = P(X \leq x) = 1 - \mathrm{e}^{-(x/a)^c} = 1 - \mathrm{e}^{-x^c/b}$$

för $x \geq 0$. Parametern $b = a^c$.

Tätheten för X fås ur

$$f_X(x) = \frac{d}{dx} F_X(x) = 0 - \mathrm{e}^{-x^c/b} \frac{-1}{b} c x^{c-1} = \frac{c x^{c-1}}{b} \mathrm{e}^{-x^c/b} = \frac{c x^{c-1}}{a^c} \mathrm{e}^{-(x/a)^c} \quad x \geq 0.$$

Vi söker medianen μ, den punkt sådan att

$$P(X \leq \mu) = P(X \geq \mu).$$

Men då är

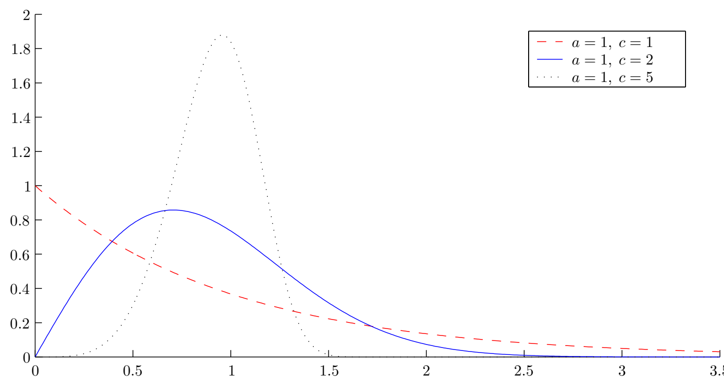
$$2P(X \leq \mu) = P(X \leq \mu) + P(X \geq \mu) = 1 + \underbrace{P(X = \mu)}_{=0} = 1$$

så μ bestäms ur

$$P(X \leq \mu) = F_X(\mu) = \frac{1}{2}.$$

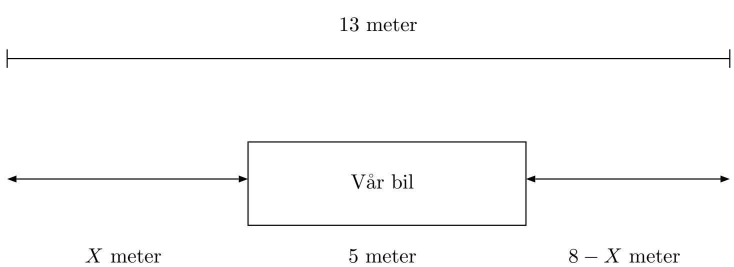
Med uttrycket för fördelningsfunktionen får man

$$\begin{aligned} 1/2 &= 1 - \mathrm{e}^{-\mu^c/b} \\ \mathrm{e}^{-\mu^c/b} &= 1/2 \\ -\mu^c/b &= \ln(1/2) \\ \mu^c/b &= \ln(2) \\ \mu &= (b\ln(2))^{1/c} = a(\ln(2))^{1/c}. \end{aligned}$$



Täthetsfunktionen $f_X(x)$ för Weibullfördelningen vid några värden på a och c.

3.15 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

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

3.16 Låt X beskriva livslängden för en transistor. Modell: X är exponentialfördelad med parameter $m = 10000$, dvs.

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

för $x \geq 0$ och $f_X(x) = 0$ om $x < 0$.

a) Vi erhåller

$$P(X < 6000) = \int_{-\infty}^{6000} f_X(x)\,dx = \int_0^{6000} \frac{1}{m}\mathrm{e}^{-x/m}\,dx = \left[-\mathrm{e}^{-x/m}\right]_0^{6000} = 1 - \mathrm{e}^{-0.6} \approx 0.4512$$

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

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

3.17 Felintensiteten $\lambda(x)$ kan uttryckas i $f_X(x)$ och $F_X(x)$ genom följande iakttagelse:

$$\lambda(x) = \lim_{h\to 0} \frac{1}{h} \cdot P\,(x < X \le X+h | X > x)$$

definition av betingning

$$= \lim_{h\to 0} \frac{1}{h} \cdot P\,(\{x < X \le X+h\} \cap \{X > x\})\,/P\,(X > x)$$

händelsen $\{x < X < x+h\} \subset \{x < X\}$

$$= \lim_{h\to 0} \frac{1}{h} \cdot P\,(x < X \le X+h)\,/(1 - P\,(X \le x))$$

uttryck sannolikheterna med hjälp av fördelningsfunktionen

$$\begin{aligned} &= \lim_{h\to 0} \frac{1}{h} \cdot (F_X(x+h) - F_X(x))/(1-F_X(x)) \\ &= \frac{1}{1-F_X(x)} \lim_{h\to 0} \frac{F_X(x+h)-F_X(x)}{h} \end{aligned}$$

identifiera definitionen av derivata

$$\begin{aligned} &= \frac{1}{1-F_X(x)} \frac{d}{dx} F_X(x) \\ &= \frac{1}{1-F_X(x)} f_X(x). \end{aligned}$$

Om X är exponentialfördelad så är $f_X(x) = \frac{1}{m}\mathrm{e}^{-x/m}$ för $x \ge 0$. Således är

$$F_X(x) = P\,(X \le x) = \int_0^x f_X(t)\,dt = \int_0^x \frac{1}{m}\mathrm{e}^{-t/m} dt = 1 - \mathrm{e}^{-x/m}$$

och

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

dvs. konstant felintensitet. (Detta definierar exponentialfördelningen.)

Om X är Weibullfördelad så är $F_X(x) = 1 - \mathrm{e}^{-(x/a)^c}$ för $x \ge 0$. Således är

$$f_X(x) = \frac{d}{dx} F_X(x) = 0 - \mathrm{e}^{-(x/a)^c} \frac{-cx^{c-1}}{a^c} = \frac{cx^{c-1}}{a^c}\,\mathrm{e}^{-(x/a)^c}.$$

och

$$\lambda(x) = \frac{f_X(x)}{1 - F_X(x)} = \frac{\frac{cx^{c-1}}{a^c}\,\mathrm{e}^{-(x/a)^c}}{1-(1-\mathrm{e}^{-(x/a)^c})} = \frac{cx^{c-1}}{a^c}.$$

Alltså är $\lambda(x) \begin{Bmatrix} \text{avtagande för } c<1 \\ \text{konstant för } c=1 \\ \text{växande för } c>1 \end{Bmatrix}$. Då $c = 1$ är X exponentialfördelad med parameter a.

3.18 Låt X vara tiden då maskin A går sönder. Modell: X är exponentialfördelad med $E(X) = m = 10$. Då är sannolikheten att

$$P(X \in (1+2k,\ 2+2k]) = \int_{1+2k}^{2+2k} f_X(x)\,dx = \int_{1+2k}^{2+2k} \frac{1}{m}\mathrm{e}^{-x/m}dx = \mathrm{e}^{-(1+2k)/m} - \mathrm{e}^{-(2+2k)/m}.$$

Sannolikheten att maskin A får sönder medan arbetaren är vid maskin B är

$$\begin{aligned}
P\left(X \in \bigcup_{k=0}^{\infty}(1+2k,\ 2+2k]\right) &= P\left(\bigcup_{k=0}^{\infty}\{X \in (1+2k,\ 2+2k]\}\right) \\
&= \sum_{k=0}^{\infty} P(X \in (1+2k,\ 2+2k]) \\
&= \sum_{k=0}^{\infty} \mathrm{e}^{-(1+2k)/m} - \mathrm{e}^{-(2+2k)/m} \\
&= \sum_{k=0}^{\infty}(\mathrm{e}^{-1/m} - \mathrm{e}^{-2/m})\mathrm{e}^{-2k/m} = (\mathrm{e}^{-1/m} - \mathrm{e}^{-2/m})\sum_{k=0}^{\infty}(\underbrace{\mathrm{e}^{-2/m}}_{=q})^k \\
&= (\mathrm{e}^{-1/m} - \mathrm{e}^{-2/m})\sum_{k=0}^{\infty} q^k = (\mathrm{e}^{-1/m} - \mathrm{e}^{-2/m})\frac{1}{1-q} \\
&= (\mathrm{e}^{-1/m} - \mathrm{e}^{-2/m})\frac{1}{1-\mathrm{e}^{-2/m}} = \frac{\mathrm{e}^{1/m}-1}{\mathrm{e}^{2/m}-1} \\
&= \frac{\mathrm{e}^{1/m}-1}{(\mathrm{e}^{1/m}-1)(\mathrm{e}^{1/m}+1)} = \frac{1}{\mathrm{e}^{1/m}+1}.
\end{aligned}$$

Notera att då $m \to 0$ så går sannolikheten mot 0 och då $m \to \infty$ så går sannolikheten mot 1/2.

Med $m = 10$ fås sannolikheten till $(\mathrm{e}^{0.10}+1)^{-1} = 0.47502$.

4.1 Då X har $f_X(x) = 1/4$ för $0 \le x \le 4$ och Y har $f_Y(y) = 1/6$ för $0 \le y \le 6$ 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 $P(X < Y)$.

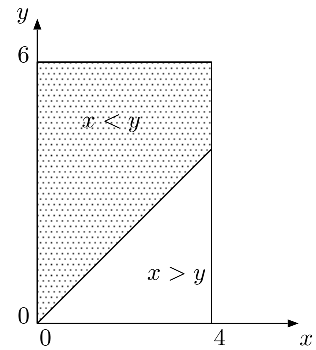
Alternativ 1: Enligt definitionen

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

Alternativ 2: Ur figuren. Under likformig fördelning att få ett utfall i det markerade området 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

$$P(X < Y) = \frac{24 - 8}{24} = \frac{2}{3}.$$



4.2 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 $f_X(x) = \frac{1}{a}\mathrm{e}^{-x/a}$ för $x \geq 0$ och Y har $f_Y(y) = \frac{1}{b}\mathrm{e}^{-y/b}$ för $y \geq 0$ och X och Y är oberoende så är

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{a}\mathrm{e}^{-x/a}\frac{1}{b}\mathrm{e}^{-y/b}$$

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

$$\begin{aligned}
P(X < Y) &= \iint\limits_{x<y} f_{X,Y}(x, y)\, dydx = \int_0^\infty \int_x^\infty f_{X,Y}(x, y)\, dydx = \int_0^\infty \int_x^\infty \frac{1}{a}\mathrm{e}^{-x/a}\frac{1}{b}\mathrm{e}^{-y/b}\, dydx \\
&= \int_0^\infty \frac{1}{a}\mathrm{e}^{-x/a}\frac{1}{b}\left[\mathrm{e}^{-y/b}(-b)\right]_x^\infty dx = \int_0^\infty \frac{1}{a}\mathrm{e}^{-x/a}\mathrm{e}^{-x/b}\, dx = \int_0^\infty \frac{1}{a}\mathrm{e}^{-x(1/a+1/b)}\, dx \\
&= \frac{1}{a}\left[\mathrm{e}^{-x(1/a+1/b)}\frac{-1}{\frac{1}{a}+\frac{1}{b}}\right]_0^\infty = \frac{1}{a}\,\frac{1}{\frac{1}{a}+\frac{1}{b}} = \frac{b}{a+b}.
\end{aligned}$$

Om $a = 2b$ är sannolikheten $b/(2b + b) = 1/3$.

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

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

för $x \geq 0$, $y \geq 0$.

a) Marginalfördelningarna bestäms ur

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

b) Då $c = 0$ är

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

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 $f_X(x)f_Y(y) \neq f_{X,Y}(x, y)$, dvs X och Y är ej oberoende.

4.4 Låt (X,Y) beskriva skillnaden mellan punktens utprickade och faktiska position där X och Y är oberoende $\mathrm{N}(0,1)$-fördelade stokastiska variabler. Avståndet mellan utprickad och faktisk position ges av $R = \sqrt{X^2+Y^2}$ och vi söker $P(R \le 2)$.

$$P(R \le 2) = P\left(\sqrt{X^2+Y^2} \le 2\right) = P\left(X^2+Y^2 \le 4\right) = \iint\limits_{(x,y):x^2+y^2\le 4} f_{X,Y}(x,y)\,dx\,dy$$

utnyttja oberoendet

$$= \iint\limits_{(x,y):x^2+y^2\le 4} f_X(x)f_Y(y)\,dx\,dy = \iint\limits_{(x,y):x^2+y^2\le 4} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2}\,dx\,dy$$

$$= \iint\limits_{(x,y):x^2+y^2\le 4} \frac{1}{2\pi}\mathrm{e}^{-(x^2+y^2)/2}\,dx\,dy$$

inför polära koordinater r och θ

$$= \iint\limits_{r^2\le 4,\theta\in[0,2\pi]} \frac{1}{2\pi}\mathrm{e}^{-r^2/2}r\,dr\,d\theta = \int_0^2 \mathrm{e}^{-r^2/2}r\,dr \underbrace{\int_0^{2\pi}\frac{1}{2\pi}\,d\theta}_{=1}$$

substitution $u = r^2/2$, $du = r\,dr$

$$= \int_0^2 \mathrm{e}^{-u}\,du \cdot 1 = 1 - \mathrm{e}^{-2}.$$

4.5 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 träffytan.

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

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

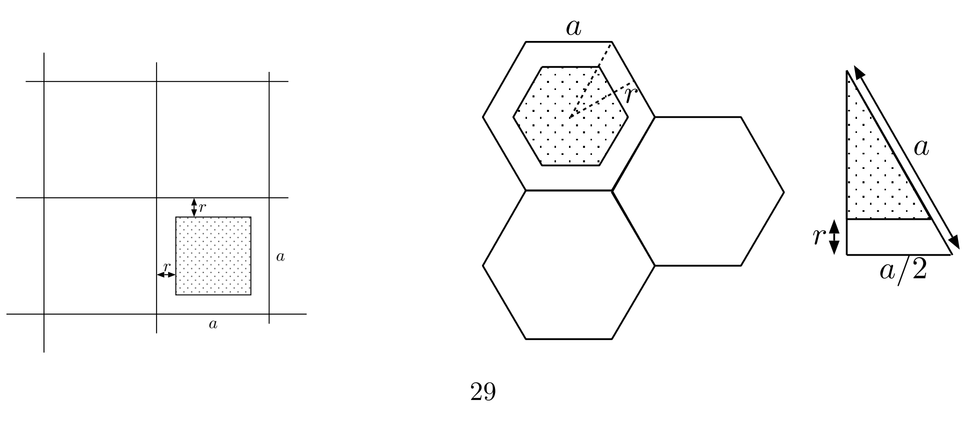
b) En hexagon med sidlängd a kan delas in i 12 trianglar enligt figuren. Triangelns höjd är (m.h.a. Pythagoras) $\sqrt{3}a/2$ så träffytan blir $\sqrt{3}a^2/8$. Ytan på avstånd större än r från triangelns bas är

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

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 $(a - 2r/\sqrt{3})^2/a^2$.)



5.1 Funktionen $g(x) = \cos((x-3)\pi)$ avbildar $x = 1, \ldots 5$, enligt

x	$\cos((x-3)\pi)$
1	1
2	-1
3	1
4	-1
5	1

dvs de möjliga värdena för $Y = g(X)$ ges av $S_Y = \{-1, 1\}$. Vi erhåller därför

$$\begin{aligned} p_Y(-1) &= P(Y=-1) = P(X=2) + P(X=4) = \frac{2}{5} \\ p_Y(1) &= P(Y=1) = P(X=1) + P(X=3) + P(X=5) = \frac{3}{5}. \end{aligned}$$

5.2 Låt U vara likformigt fördelad på $[0, 1]$ Då är $f_U(x) = 1$ då $0 \leq x \leq 1$ och

$$F_U(t) = P(U \leq t) = \int_0^t f_X(x)\,dx = \int_0^t 1\,dx = t$$

för $0 \leq t \leq 1$.

a) Med $X = a + (b-a)U$ så ges de möjliga värdena för X av $S_X = [a,\ b]$ och X har fördelningsfunktion

$$F_X(t) = P(X \leq t) = P(a + (b-a)U \leq t) = P\left(U \leq \frac{t-a}{b-a}\right) = F_U\left(\frac{t-a}{b-a}\right) = \frac{t-a}{b-a}$$

för $a \leq t \leq b$. Deriveras denna fås tätheten

$$f_X(t) = \frac{d}{dt}F_X(t) = \frac{1}{b-a}$$

dvs likformig fördelning på $[a,\ b]$.

Med $X = -m\ln(U)$ så ges de möjliga värdena för X av $S_X = [0,\ \infty)$ och X har fördelningsfunktionen

$$\begin{aligned} F_X(t) &= P(X \leq t) = P(-m\ln(U) \leq t) = P(\ln(U) \geq -t/m) = P\left(U \geq \mathrm{e}^{-t/m}\right) \\ &= 1 - P\left(U \leq \mathrm{e}^{-t/m}\right) = 1 - F_U(\mathrm{e}^{-t/m}) = 1 - \mathrm{e}^{-t/m}. \end{aligned}$$

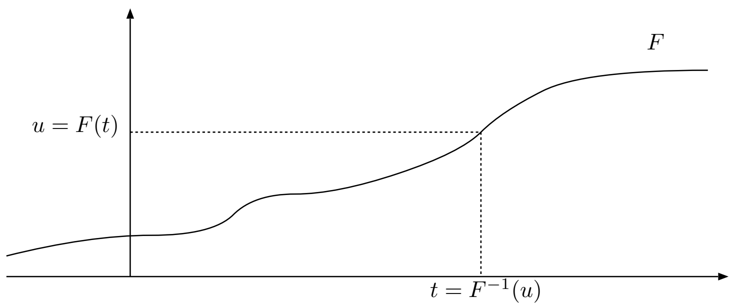
för $t \geq 0$. Deriveras denna fås tätheten

$$f_X(t) = \frac{d}{dt}F_X(t) = \frac{1}{m}\mathrm{e}^{-t/m}, \quad t \geq 0,$$

dvs exponentialfördelningen med parameter m.

b) Avbildningen $t \mapsto F(t)$ är en avbildning $(-\infty, \infty)$ till $[0, 1]$. Inversen F^{-1} avbildar $[0, 1]$ på $(-\infty, \infty)$. Med $X = F^{-1}(U)$ så har X fördelningsfunktion:

$$F_X(t) = P(X \leq t) = P\left(F^{-1}(U) \leq t\right) = \{\text{Kolla figur.}\} = P(U \leq F(t)) = F_U(F(t)) = F(t).$$



Samband mellan funktionen $F(t)$ och inversen $F^{-1}(u)$.

5.3 Låt X beskriva positionen för en jordbävnings epicentrum i förkastningssprickan. Modell: X är $\mathrm{R}[-a, a]$, dvs

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

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

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

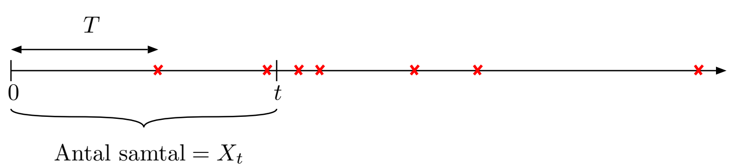
De möjliga värdena på Y är $S_Y = [d, \sqrt{a^2 + d^2}]$. För ett $t \in S_Y$ så är

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

Täthetsfunktionen fås genom derivering

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

5.4 Låt X_t beskriva antalet samtal under intervallet $[0, t]$. Låt T beskriva tiden till första samtalet.



De möjliga värdena på T är $S_T = [0, \infty)$. Notera att $\{T > t\}$ och $\{X_t = 0\}$ är samma händelse. Så, med modellen att X_t är $\mathrm{Po}(\lambda t)$ så T har fördelningsfunktion

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - P(X_t = 0) = 1 - \frac{(\lambda t)^0}{0!}\mathrm{e}^{-\lambda t} = 1 - \mathrm{e}^{-\lambda t}$$

för $t \geq 0$. Deriveras denna fås täthetsfunktionen

$$f_T(t) = \frac{d}{dt} F_T(t) = \lambda \mathrm{e}^{-\lambda t} = \frac{1}{m}\mathrm{e}^{-t/m}$$

för $t \geq 0$ och $m = 1/\lambda$. Alltså, T är exponentialfördelad med väntevärde $1/\lambda$.

5.5 Låt X beskriva den tid som fru Svensson parkerar. Modell X är likformigt fördelad på intervallet $[20, 60]$, dvs

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

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

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

5.6 Om X är kontinuerlig och har möjliga värden på $(-\infty, \infty)$ så har $Y = |X|$ möjliga värden på $S_Y = [0, \infty)$. Där har Y fördelningsfunktion

$$F_Y(t) = P(Y \le t) = P(|X| \le t) = P(-t \le X \le t) = P(-t < X \le t) = F_X(t) - F_X(-t).$$

Deriveras fördelningsfunktionen fås tätheten

$$f_Y(t) = \frac{d}{dt}\left[F_X(t) - F_X(-t)\right] = f_X(t) - f_X(-t)(-1) = f_X(t) + f_X(-t).$$

för $t \ge 0$. Om X har täthet

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-t^2/2\sigma^2}.$$

så har Y täthet

$$f_Y(t) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-t^2/2\sigma^2} + \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(-t)^2/2\sigma^2} = \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-t^2/2\sigma^2}.$$

5.7 Låt θ vara likformigt fördelad på intervallet $[-\frac{\pi}{2}, \frac{\pi}{2}]$, dvs har täthet $f_\theta(x) = 1/\pi$ för $-\frac{\pi}{2} \le x \le \frac{\pi}{2}$. Då är fördelningsfunktionen för θ

$$F_\theta(t) = P(\theta \le t) = \int_{-\infty}^{t} f_\theta(x)\,dx = \int_{-\pi/2}^{t} \frac{1}{\pi}\,dx = \frac{t + \pi/2}{\pi} \quad -\tfrac{\pi}{2} \le t \le \tfrac{\pi}{2}.$$

Träffpunkten läge på y-axeln ges av $Y = d\tan\theta$. De möjliga värdena på Y ges av $S_Y = (-\infty, \infty)$ och Y har fördelningsfunktion

$$\begin{aligned}
F_Y(t) &= P(Y \le t) = P(d\tan(\theta) \le t) = P\big(\theta \le \tan^{-1}(t/d)\big) = F_\theta(\tan^{-1}(t/d)) \\
&= \frac{1}{\pi}\tan^{-1}(t/d) + \frac{1}{2}.
\end{aligned}$$

Deriveras denna fås tätheten

$$f_Y(t) = \frac{d}{dt}F_Y(t) = \frac{1}{\pi}\,\frac{1}{1 + (t/d)^2}\,\frac{1}{d} = \frac{1}{\pi}\,\frac{d}{d^2 + t^2}.$$

5.8 Låt $Y = \min(X_1, \ldots, X_n)$ och $Z = \max(X_1, \ldots, X_n)$. Lösningen till problemet bygger på iakttagelsen att om $\max(X_1, \ldots, X_n) \le t$ så är *alla* $X_1, \ldots, X_n$ högst t. På samma sätt att om $\min(X_1, \ldots, X_n) > t$ så är *alla* $X_1, \ldots, X_n$ större än t.

Med Z enligt ovan så har Z fördelningsfunktion

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

På samma sätt har Y fördelningsfunktion

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

5.9 Låt $Y = \min(X_1,\ldots,X_n)$. Lösningen till problemet bygger på iakttagelsen att

$$\min(X_1,\ldots,X_n) > t \quad \text{är ekvivalent med} \quad X_1 > t,\ldots,X_n > t.$$

För en exponentialfördelning så är

$$P(X > t) = \int_t^\infty f_X(x)\,dx = \int_t^\infty \frac{1}{m}\mathrm{e}^{-x/m}\,dx = \Big[-\mathrm{e}^{-x/m}\Big]_t^\infty = \mathrm{e}^{-t/m}$$

för $t \ge 0$.

Alltså är

$$\begin{aligned} P(Y > t) &= P(\min(X_1,\ldots,X_n) > t) = P(X_1 > t,\ldots,X_n > t) = \{\text{oberoende}\} \\ &= P(X_1 > t)\cdots P(X_n > t) = \mathrm{e}^{-t/m}\cdots\mathrm{e}^{-t/m} = \mathrm{e}^{-tn/m}. \end{aligned}$$

Detta ger att Y har fördelningsfunktion

$$F_Y(t) = P(Y \le t) = 1 - P(Y > t) = 1 - \mathrm{e}^{-tn/m}$$

för $t \ge 0$. Med $m' = m/n$ så är

$$F_Y(t) = 1 - \mathrm{e}^{-t/m'}$$

och

$$f_Y(t) = \frac{d}{dt}F_Y(t) = \frac{1}{m'}\mathrm{e}^{-t/m'}$$

då $t \ge 0$, dvs. Y är exponentialfördelad med parameter $m' = m/n$.

5.10 Under likformig fördelning är sannolikheten att få ett utfall med radie högst r, kvoten mellan arean med högst t, dvs. cirkelskivan med radie t, och den totala arean, cirkelskivan med radie r. Alltså, med X som avstånd till mittpunkten är

$$F_X(t) = \frac{\pi t^2}{\pi r^2} = (t/r)^2$$

för $0 \le t \le r$.

Låt Y vara avståndet för en annan punkt vald på måfå oberoende av X. Då, med $Z = \min(X, Y)$, är

$$\begin{aligned} F_Z(t) &= P(Z \le t) = P(\min(X,Y) \le t) = 1 - P(\min(X,Y) > t) = 1 - P(X > t, Y > t) \\ &= 1 - P(X > t)\,P(Y > t) = 1 - (1 - P(X \le t))(1 - P(Y \le t)) \\ &= 1 - (1 - F_X(t))(1 - F_Y(t)) = 1 - (1 - (t/r)^2)^2, \end{aligned}$$

för $0 \le t \le r$.

5.11 Låt X_1 och X_2 beskriva oberoende belastningsgränserna för tråd 1 resp. tråd 2 i wiren. Wiren går sönder vid belastning a om $\min(X_1, X_2) < a/2$ och $\max(X_1, X_2) < a$.

Alternativ 1: Wiren går sönder med sannolikhet

$$
\begin{aligned}
&P(\{\min(X_1,X_2) < a/2\} \cap \{\max(X_1,X_2) < a\}) \\
&\quad = \{\textit{utnyttja } P(A\cap B) = P(B) - P(B \cap A^\star)\} \\
&\quad = P(\max(X_1,X_2) < a) - P(\{\max(X_1,X_2) < a\} \cap \{\min(X_1,X_2) \geq a/2\}) \\
&\quad = \left\{ \begin{array}{l} \textit{utnyttja } \max(a,b) < t \Leftrightarrow a < t, b < t \textit{ och} \\ \min(a,b) > t \Leftrightarrow a > t, b > t \end{array} \right\} \\
&\quad = P(X_1 < a, X_2 < a) - P(a/2 \leq X_1 < a,\ a/2 \leq X_2 < a) \\
&\quad = \{\textit{utnyttja oberoendet}\} \\
&\quad = P(X_1 < a)\,P(X_2 < a) - P(a/2 \leq X_1 < a)\,P(a/2 \leq X_2 < a) \\
&\quad = F_X(a)^2 - (F_X(a) - F_X(a/2))^2 \\
&\quad = 2F_X(a)F_X(a/2) - F_X(a/2)^2.
\end{aligned}
$$

Alternativ 2: Låt p_1 vara sannolikheten att en tråd har en belastningsgräns på intervallet $[0, a/2]$ och p_2 att den har det på intervallet $(a/2, a]$. Då är $p_1 = F_X(a/2)$ och $p_2 = F_X(a) - F_X(a/2)$.

Wiren går sönder om båda trådarna ligger på intervallet $[0, a/2]$ (sker med sannolikhet p_1^2) eller om den ena ligger på intervallet $[0, a/2]$ och den andra på $(a/2, a]$ (sker med sannolikhet $2p_1p_2$). Sannolikheten för att wiren går sönder blir alltså

$$
p_1^2 + 2p_1p_2 = F_X(a/2)^2 + 2(F_X(a) - F_X(a/2))F_X(a/2) = 2F_X(a)F_X(a/2) - F_X(a/2)^2.
$$

5.12 En punkt vald på måfå i intervallet $[0,1]$ delar intervallet i två bitar av längd U och $1-U$ där U är likformigt fördelad på $[0,1]$ och av symmetriskäl, även $1-U$.

a) Med $X = \min(U, 1-U)$ så ges de möjliga värdena på X av $S_X = [0, 1/2]$ och

$$
\begin{aligned}
P(X > t) &= P(\min(U, 1-U) > t) = P(U > t, 1-U > t)\,P(t < U < 1-t) \\
&= (1-t) - t = 1 - 2t
\end{aligned}
$$

för $0 \leq t \leq 1/2$. Alltså är $F_X(t) = P(X \leq t) = 1 - P(X > t) = 2t$ och har tätheten $f_X(t) = \frac{d}{dt}F_X(t) = 2$ då $0 \leq t \leq 1/2$. Alltså, X är likformigt fördelad på intervallet $[0, 1/2]$.

b) Låt $Y = \max(U, 1-U)$ då är med X enligt är de möjliga värdena på kvoten Y/X $[1, \infty)$ och fördelningsfunktionen för kvoten ges av

$$
\begin{aligned}
F_{Y/X}(t) &= P(Y/X \leq t) = P(Y \leq tX) = P\left(Y \leq tX,\ U < \tfrac{1}{2}\right) + P\left(Y \leq tX,\ U \geq \tfrac{1}{2}\right) \\
&= P\left(1 - U \leq tU,\ U < \tfrac{1}{2}\right) + P\left(U \leq t(1-U),\ U \geq \tfrac{1}{2}\right) \\
&= P\left(1/(t+1) \leq U < \tfrac{1}{2}\right) + P\left(\tfrac{1}{2} \leq U \leq t/(1+t)\right) \\
&= \left(\frac{1}{2} - \frac{1}{t+1}\right) + \left(\frac{t}{1+t} - \frac{1}{2}\right) = \frac{t-1}{t+1},
\end{aligned}
$$

för $t \geq 1$.

5.13 Låt X_1 och X_2 beskriva livslängden för elektronrören. Modell: X_1 och X_2 är oberoende och exponentialfördelade med parameter $m > 0$.

Alternativ 1: Med $T = X_1 + X_2$ så har T fördelningsfunktion

$$\begin{aligned}
F_T(t) &= P(X \le t) = P(X_1 + X_2 \le t) = \iint\limits_{(x,y):x+y\le t} f_{X_1,X_2}(x,y)\,dxdy \\
&= \int_0^t \int_0^{t-y} f_{X_1}(x) f_{X_2}(y)\,dxdy = \int_0^t \int_0^{t-y} \frac{1}{m}\mathrm{e}^{-x/m}\frac{1}{m}\mathrm{e}^{-y/m}\,dxdy \\
&= \frac{1}{m^2}\int_0^t \mathrm{e}^{-y/m}\left[\mathrm{e}^{-x/m}(-m)\right]_0^{t-y} dy = \frac{1}{m}\int_0^t \mathrm{e}^{-y/m} - \mathrm{e}^{-t/m} dy \\
&= \frac{1}{m}\left[\mathrm{e}^{-y/m}(-m) - \mathrm{e}^{-t/m} y\right]_0^t = 1 - \mathrm{e}^{-t/m} - \mathrm{e}^{-t/m}\frac{t}{m}
\end{aligned}$$

för $t \ge 0$.

Alternativ 2: Med faltningsformeln får T tätheten

$$\begin{aligned}
f_T(t) &= \int_0^t f_{X_1}(x) f_{X_2}(t-x)\,dx = \int_0^t \frac{1}{m}\mathrm{e}^{-x/m}\frac{1}{m}\mathrm{e}^{-(t-x)/m}\,dx = \frac{1}{m^2}\int_0^t \mathrm{e}^{-t/m}\,dx \\
&= \frac{t}{m^2}\mathrm{e}^{-t/m}
\end{aligned}$$

för $t \ge 0$. Nu är

$$\begin{aligned}
P(T > t) &= \int_t^\infty f_T(x)\,dx = \int_t^\infty \frac{x}{m^2}\mathrm{e}^{-x/m}\,dx = \left[\frac{-x}{m}\mathrm{e}^{-x/m}\right]_t^\infty + \int_t^\infty \frac{1}{m}\mathrm{e}^{-x/m}\,dx \\
&= \frac{t}{m}\mathrm{e}^{-t/m} + \left[-\mathrm{e}^{-x/m}\right]_t^\infty = \frac{t}{m}\mathrm{e}^{-t/m} + \mathrm{e}^{-t/m}
\end{aligned}$$

och

$$F_T(t) = P(T \le t) = 1 - P(T > t) = 1 - \frac{t}{m}\mathrm{e}^{-t/m} - \mathrm{e}^{-t/m}$$

för $t \ge 0$.

Vi söker

$$\begin{aligned}
P(T > 400) &= 1 - P(T \le 400) = 1 - F_T(400) = \mathrm{e}^{-400/m} + \mathrm{e}^{-400/m}\frac{400}{m} = 3\mathrm{e}^{-2} \\
&= 0.40601.
\end{aligned}$$

5.14 Med X och Y som oberoende Poissonfördelade stokastiska variabler med parametrar m_x och m_y. Vi skriver händelsen

$$\{X + Y = n\} = \bigcup_k \{X = k, Y = n-k\}$$

så sannolikheten

$$\begin{aligned}
P(X+Y=n) &= P\left(\bigcup_k \{X = k, Y = n-k\}\right) = \sum_k P(X = k, Y = n-k) \\
&= \sum_k P(X = k)\, P(Y = n-k) = \sum_{k=0}^n \frac{m_x^k}{k!}\mathrm{e}^{-m_x}\frac{m_y^{n-k}}{(n-k)!}\mathrm{e}^{-m_y} \\
&= \sum_{k=0}^n \frac{m_x^k}{k!}\frac{m_y^{n-k}}{(n-k)!}\frac{n!}{n!}\,\mathrm{e}^{-(m_x+m_y)} = \frac{1}{n!}\mathrm{e}^{-(m_x+m_y)}\underbrace{\sum_{k=0}^n \binom{n}{k} m_x^k m_y^{n-k}}_{=(m_x+m_y)^n} \\
&= \frac{(m_x+m_y)^n}{n!}\mathrm{e}^{-(m_x+m_y)} = \frac{m^n}{n!}\mathrm{e}^{-m}.
\end{aligned}$$

där $m = m_x + m_y$, för $n = 0, 1, 2, \ldots,$. Dvs, $X + Y$ är Poissonfördelad med parameter $m = m_x + m_y$.

5.15 Låt $X_1, X_2, \ldots, X_n$ vara oberoende exponentialfördelade stokastiska variabler med parameter m. Låt $Y_n = \sum_{i=1}^{n} X_i$.

a) Med induktion visar vi att Y_n följer en $\Gamma(n, m)$-fördelning.

Basfall: Med $n = 1$ får vi att $Y_1 = X_1$ är exponentialfördelad, dvs har tätheten

$$f_{Y_1}(x) = f_{X_1}(x) = \frac{1}{m}\mathrm{e}^{-x/m} = \frac{1}{m^n\Gamma(n)}x^{n-1}\mathrm{e}^{-x/m}$$

tätheten för en $\Gamma(n, m)$-fördelning, $n = 1$.

Induktionssteg. Antag att Y_n har $\Gamma(n, m)$-fördelning. Visa att detta leder till att Y_{n+1} har en $\Gamma(n + 1, m)$-fördelning.

$$\begin{aligned} f_{Y_{n+1}}(t) &= \int_0^t f_{Y_n}(s) f_X(t-s)\,ds = \int_0^t \frac{1}{m^n\Gamma(n)}s^{n-1}\mathrm{e}^{-s/m}\,\frac{1}{m}\mathrm{e}^{-(t-s)/m}\,ds \\ &= \frac{1}{m^{n+1}\Gamma(n)}\mathrm{e}^{-t/m}\int_0^t s^{n-1}\,ds = \frac{1}{m^{n+1}\Gamma(n)}\mathrm{e}^{-t/m}\frac{t^n}{n} = \frac{1}{m^{n+1}\underbrace{n\Gamma(n)}_{=\Gamma(n+1)}}t^n\mathrm{e}^{-t/m} \\ &= \frac{1}{m^{n+1}\Gamma(n+1)}t^n\mathrm{e}^{-t/m}, \end{aligned}$$

för $t \geq 0$, dvs tätheten för en $\Gamma(n + 1, m)$-fördelning.

b) Låt X och Y vara två oberoende $\Gamma(n_1, m)$- och $\Gamma(n_2, m)$-fördelade stokastiska variabler.

Alternativ 1: Via representationen av $\Gamma(\,\cdot\,, m)$-fördelningen i a). Om n_1 och n_2 är positivt heltaliga så kan enligt a) X och Y skrivas

$$X = \sum_{i=1}^{n_1} X_i \qquad Y = \sum_{i=1}^{n_2} Y_i$$

där $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ är oberoende exponentialfördelade med parameter m. Men då är

$$X + Y = \sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i$$

en summa av $(n_1 + n_2)$ oberoende exponentialfördelade stokastiska variabler med parameter m och alltså är $\Gamma(n_1 + n_2, m)$-fördelad.

Alternativ 2: Via faltningsformeln.

$$\begin{aligned} f_{X+Y}(t) &= \int_0^t f_X(s) f_Y(t-s)\,ds \\ &= \int_0^t \frac{1}{m^{n_1}\Gamma(n_1)}s^{n_1-1}\mathrm{e}^{-s/m}\,\frac{1}{m^{n_2}\Gamma(n_2)}(t-s)^{n_2-1}\mathrm{e}^{-(t-s)/m}\,ds \\ &= \frac{1}{m^{n_1+n_2}\Gamma(n_1)\Gamma(n_2)}\mathrm{e}^{-t/m}\int_0^t s^{n_1-1}(t-s)^{n_2-1}\,ds \end{aligned}$$

Partialintegrering ger

$$\begin{aligned}
\int_0^t s^{n_1-1}(t-s)^{n_2-1}\,ds &= \underbrace{\left[s^{n_1-1}(t-s)^{n_2}\frac{-1}{n_2}\right]_0^t}_{=0} + \frac{n_1-1}{n_2}\int_0^t s^{n_1-2}(t-s)^{n_2}\,ds \\
&= 0 + \frac{n_1-1}{n_2}\,\frac{n_1-2}{n_2+1}\int_0^t s^{n_1-3}(t-s)^{n_2+1}\,ds = \cdots \\
&= \frac{(n_1-1)(n_1-2)\cdots 1}{n2(n_2+1)\cdots(n_2+n_1-1)}\,t^{n_1+n_2-1} \\
&= \frac{(n_1-1)!(n_2-1)!}{(n_2+n_1-1)!}\,t^{n_1+n_2-1} \\
&= \frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n_1+n_2)}\,t^{n_1+n_2-1}
\end{aligned}$$

Så tätheten är

$$\begin{aligned}
f_{X+Y}(t) &= \frac{1}{m^{n_1+n_2}\Gamma(n_1)\Gamma(n_2)}\mathrm{e}^{-t/m}\,\frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n_1+n_2)}\,t^{n_1+n_2-1} \\
&= \frac{1}{m^{n_1+n_2}\Gamma(n_1+n_2)}t^{n_1+n_2-1}\mathrm{e}^{-t/m}
\end{aligned}$$

för $t > 0$ är tätheten för en $\Gamma(n_1+n_2, m)$-fördelning.

5.16 Låt X och Y vara oberoende och $\mathrm{N}(0,1)$. I polära koordinater (R,Θ) är

$$(X,Y) = (R\cos(\Theta), R\sin(\Theta)).$$

$$\begin{aligned}
P\left(R \le s, \Theta \le t\right) &= P\left(\sqrt{X^2+Y^2} \le s, \tan^{-1}(Y/X) \le t\right) \\
&= P\left(X^2+Y^2 \le s^2, \tan^{-1}(Y/X) \le t\right) = P\left((X,Y) \in D\right) \\
&= \iint\limits_D f_{X,Y}(x,y)\,dxdy = \iint\limits_D f_X(x)f_Y(y)\,dxdy \\
&= \iint\limits_D \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\,\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2}\,dxdy = \iint\limits_D \frac{1}{2\pi}\mathrm{e}^{-(x^2+y^2)/2}\,dxdy \\
&= \iint\limits_D \underbrace{\frac{1}{2\pi}\mathrm{e}^{-r^2/2}r}_{=f_{R,\Theta}(r,\theta)}\,drd\theta = \int_0^t\int_0^s f_{R,\Theta}(r,\theta)\,drd\theta
\end{aligned}$$

Alltså: $f_{R,\Theta}(r,\theta) = \frac{1}{2\pi}\mathrm{e}^{-r^2/2}r$. Vidare fås fördelningsfunktionerna

$$\begin{aligned}
F_R(s) &= P\left(R \le s\right) = P\left(R \le s, \Theta \le 2\pi\right) = \int_0^{2\pi}\int_0^s \frac{1}{2\pi}\mathrm{e}^{-r^2/2}r\,drd\theta \\
&= \int_0^{2\pi}\int_0^{s^2/2}\frac{1}{2\pi}\mathrm{e}^{-u}\,dud\theta = \int_0^{2\pi}\left[\frac{1}{2\pi}\mathrm{e}^{-u}(-1)\right]_0^{s^2/2}d\theta \\
&= \int_0^{2\pi}\frac{1}{2\pi}\left[1-\mathrm{e}^{-s^2/2}\right]d\theta = 1-\mathrm{e}^{-s^2/2}
\end{aligned}$$

så $f_R(s) = \frac{d}{ds}F_R(s) = 0 - \mathrm{e}^{-s^2/2}(-s) = s\mathrm{e}^{-s^2/2}$. På samma sätt

$$\begin{aligned}
F_\Theta(t) &= P(\Theta \le t) = P(R \le \infty, \Theta \le t) = \int_0^t \int_0^\infty \frac{1}{2\pi}\mathrm{e}^{-r^2/2} r\, drd\theta \\
&= \int_0^t \int_0^\infty \frac{1}{2\pi}\mathrm{e}^{-u}\, dud\theta = \int_0^t \left[\frac{1}{2\pi}\mathrm{e}^{-u}(-1)\right]_0^\infty d\theta \\
&= \int_0^t \frac{1}{2\pi} d\theta = \frac{t}{2\pi}
\end{aligned}$$

så $f_\Theta(t) = \frac{d}{dt}F_\Theta(t) = \frac{1}{2\pi}$. Alltså är

$$f_{R,\Theta}(s,t) = \frac{1}{2\pi}\mathrm{e}^{-s^2/2} s = f_\Theta(t)\, f_R(s)$$

för alla s och t, vilket säger att R och Θ är oberoende.

5.17 Låt X_1 och X_2 vara oberoende och exponentialfördelade med parameter $m = 1$. Då är tätheten $f_X(x) = \mathrm{e}^{-x}$, för $x \ge 0$.

Med $Y = X_1/X_2$ så ges de möjliga värdena för Y av $S_Y = [0, \infty)$ och Y har fördelningsfunktion

$$\begin{aligned}
F_Y(t) &= P(Y \le t) = P(X_1/X_2 \le t) = P(X_1 \le tX_2) = \iint_{(x,y):x \le ty} f_{X_1,X_2}(x,y)\, dxdy \\
&= \int_0^\infty \int_0^{ty} f_{X_1}(x) f_{X_2}(y)\, dxdy = \int_0^\infty \int_0^{ty} \mathrm{e}^{-x}\mathrm{e}^{-y}\, dxdy = \int_0^\infty \mathrm{e}^{-y}\left[1 - \mathrm{e}^{-ty}\right] dy \\
&= \left[\mathrm{e}^{-y}(-1) + \mathrm{e}^{-(t+1)y}(t+1)\right]_0^\infty = 1 - \frac{1}{t+1} = \frac{t}{t+1}.
\end{aligned}$$

Således är tätheten för Y

$$f_Y(t) = \frac{d}{dt}F_Y(t) = \frac{1}{t+1} - \frac{t}{(t+1)^2} = \frac{1}{(t+1)^2}, \quad t \ge 0.$$

Låt nu $Y_1 = X_1 + X_2$ och $Y_2 = X_1/(X_1 + X_2)$. Då har Y_1 möjliga värden $[0, \infty)$ och Y_2 möjliga värden $[0, 1]$. För ett $s \ge 0$ och $t \in [0, 1]$ är

$$P(Y_1 \le s, Y_2 \le t) = P\left(X_1 + X_2 \le s, \frac{X_1}{X_1 + X_2} \le t\right) = P\left(\frac{1-t}{t}X_1 \le X_2 \le s - X_1\right)$$

Intervallet $[\frac{1-t}{t}X_1, s - X_1]$ innebär att

$$\left(\frac{1-t}{t} + 1\right) X_1 \le s \quad \Rightarrow \quad X_1 \le st$$

så

$$\begin{aligned}
P(Y_1 \le s, Y_2 \le t) &= P\left(\frac{1-t}{t}X_1 \le X_2 \le s - X_1\right) = \int_0^{ts} \int_{(1-t)y/t}^{s-y} f_{X_1}(y) f_{X_2}(x)\, dxdy \\
&= \int_0^{ts} \int_{(1-t)y/t}^{s-y} \mathrm{e}^{-y}\mathrm{e}^{-x}\, dxdy = \int_0^{ts} \mathrm{e}^{-y}\left[\mathrm{e}^{-x}(-1)\right]_{(1-t)y/t}^{s-y} dy \\
&= \int_0^{ts} \mathrm{e}^{-y/t} - \mathrm{e}^{-s} dy = \left[\mathrm{e}^{-y/t}(-t) - \mathrm{e}^{-s}y\right]_0^{ts} = t\left(1 - \mathrm{e}^{-s}(1+s)\right)
\end{aligned}$$

för $s \ge 0$ och $t \in [0, 1]$.

Nu är

$$F_{Y_1}(s) \;=\; P(Y_1 \le s) = P(Y_1 \le s, Y_2 \le 1) = 1 - \mathrm{e}^{-s}(1+s)$$

så

$$f_{Y_1}(s) = \frac{d}{ds} F_{Y_1}(s) = s\mathrm{e}^{-s}$$

för $s \ge 0$. Vidare,

$$F_{Y_2}(t) \;=\; P(Y_2 \le t) = P(Y_1 \le \infty, Y_2 \le t) = t$$

så

$$f_{Y_2}(t) = \frac{d}{dt} F_{Y_2}(t) = 1$$

för $0 \le t \le 1$. Vi ser även att den simultana tätheten

$$f_{Y_1,Y_2}(s,t) = \frac{\partial}{\partial s}\frac{\partial}{\partial t} P(Y_1 \le s, Y_2 \le t) = s\mathrm{e}^{-s} = f_{Y_1}(s) f_{Y_2}(t)$$

så Y_1 och Y_2 är oberoende.

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

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

Variansen kan beräknas på två sätt:

Alternativ 1: Ur definitionen

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

Alternativ 2: Via

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

och sedan

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

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

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

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

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

Alternativ 3: Via lineariteten för väntevärden.

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

och

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

6.2 Mätinstrumentet ger ett mätfel X med täthet $f_X(x) = 100(1-100|x|)$ för $-0.01 \le x \le 0.01$. 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

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

så

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

6.3 Låt X vara en kontinuerlig stokastisk variabel. Dess täthetsfunktion fås genom att derivera fördelningsfunktionen, dvs

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{20 - x}{200}, \quad \text{för } 0 \leq x \leq 20,$$

och $f_X(x) = 0$ för övrigt. Vi erhåller väntevärdet enligt

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx = \int_0^{20} x \, \frac{20 - x}{200}\, dx = \frac{1}{200} \left[20\frac{x^2}{2} - \frac{x^3}{3} \right]_0^{20} = \frac{20}{3}.$$

Vi beräknar härnäst $E\left(X^2\right)$.

$$E\left(X^2\right) \;=\; \int_{-\infty}^{\infty} x^2\, f_X(x)\, dx = \int_0^{20} x^2\, \frac{20 - x}{200}\, dx = \frac{1}{200} \left[20\frac{x^3}{3} - \frac{x^4}{4} \right]_0^{20} = \frac{200}{3}.$$

Det ger oss variansen

$$V(X) = E\left(X^2\right) - (E(X))^2 = \frac{200}{3} - \left(\frac{20}{3}\right)^2 = \frac{200}{9}$$

och standardavvikelse

$$D(X) = \sqrt{V(X)} = \frac{10\sqrt{2}}{3} \approx 4.71.$$

6.4 X är exponentialfördelad med parameter $m > 0$, dvs

$$f_X(x) = \frac{1}{m} \mathrm{e}^{-x/m}, \quad x > 0.$$

Vi får då att

$$\begin{aligned} E(X) &= \int x f_X(x)\, dx = \int_0^{\infty} x\, \frac{1}{m} \mathrm{e}^{-x/m}\, dx = \underbrace{\left[x\, \frac{1}{m} \mathrm{e}^{-x/m} (-m) \right]_0^{\infty}}_{=0} + \int_0^{\infty} \mathrm{e}^{-x/m}\, dx \\ &= \left[\mathrm{e}^{-x/m} (-m) \right]_0^{\infty} = m. \end{aligned}$$

Vi kan bestämma $V(X)$ på två sätt.

Alternativ 1:

$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) = \int (x - m)^2\, f_X(x)\, dx = \int_0^{\infty} (x - m)^2\, \frac{1}{m} \mathrm{e}^{-x/m}\, dx \\ &= \underbrace{\left[(x - m)^2\, \frac{1}{m} \mathrm{e}^{-x/m} (-m) \right]_0^{\infty}}_{=m^2} + \int_0^{\infty} 2(x - m)\, \mathrm{e}^{-x/m}\, dx \\ &= m^2 + \underbrace{\left[2(x - m) \mathrm{e}^{-x/m} (-m) \right]_0^{\infty}}_{=-2m^2} + \int_0^{\infty} 2\mathrm{e}^{-x/m} m\, dx \\ &= -m^2 + \underbrace{\left[2m\mathrm{e}^{-x/m} (-m) \right]_0^{\infty}}_{=2m^2} = m^2. \end{aligned}$$

Alternativ 2:

$$
\begin{aligned}
E\left(X^2\right) &= \int x^2\, f_X(x)\, dx = \int_0^\infty x^2\, \frac{1}{m} \mathrm{e}^{-x/m}\, dx \\
&= \underbrace{\left[x^2\, \frac{1}{m} \mathrm{e}^{-x/m}(-m)\right]_0^\infty}_{=0} + \int_0^\infty 2x\mathrm{e}^{-x/m}\, dx \\
&= \underbrace{\left[2x\mathrm{e}^{-x/m}(-m)\right]_0^\infty}_{=0} + 2m \int_0^\infty \mathrm{e}^{-x/m}\, dx = \left[2m\mathrm{e}^{-x/m}(-m)\right]_0^\infty = 2m^2.
\end{aligned}
$$

Vidare så är

$$V(X) = E\left(X^2\right) - (E(X))^2 = 2m^2 - m^2 = m^2.$$

Alltså är

$$D(X) = \sqrt{V(X)} = \sqrt{m^2} = m.$$

6.5 X är Po(m) dvs.

$$P(X = k) = \frac{m^k}{k!}\mathrm{e}^{-m}, \quad k = 0, 1, 2, \ldots$$

Nu är

$$
\begin{aligned}
E(X) &= \sum_k kP(X = k) = \sum_{k=0}^\infty k\, \frac{m^k}{k!}\mathrm{e}^{-m} = \sum_{k=1}^\infty k\, \frac{m^k}{k!}\mathrm{e}^{-m} = \sum_{k=1}^\infty \frac{m^k}{(k-1)!}\mathrm{e}^{-m} \\
&= m\sum_{k=1}^\infty \frac{m^{k-1}}{(k-1)!}\mathrm{e}^{-m} = m\sum_{k=0}^\infty \frac{m^k}{k!}\mathrm{e}^{-m} = m\underbrace{\sum_{k=0}^\infty P(X = k)}_{=1} = m.
\end{aligned}
$$

Vidare så

$$
\begin{aligned}
E\left(X^2\right) &= E(X(X-1) + X) = \sum_k (k(k-1) + k)P(X = k) \\
&= \sum_k k(k-1)P(X = k) + \underbrace{\sum_k kP(X = k)}_{=E(X)=m} = m + \sum_{k=0}^\infty k(k-1)\, \frac{m^k}{k!}\mathrm{e}^{-m} \\
&= m + \sum_{k=2}^\infty k(k-1)\, \frac{m^k}{k!}\mathrm{e}^{-m} = m + \sum_{k=2}^\infty \frac{m^k}{(k-2)!}\mathrm{e}^{-m} = m + m^2\sum_{k=2}^\infty \frac{m^{k-2}}{(k-2)!}\mathrm{e}^{-m} \\
&= m + m^2\sum_{k=0}^\infty \frac{m^k}{k!}\mathrm{e}^{-m} = m + m^2\underbrace{\sum_{k=0}^\infty P(X = k)}_{=1} = m + m^2.
\end{aligned}
$$

Alltså är

$$V(X) = E\left(X^2\right) - (E(X))^2 = (m + m^2) - m^2 = m$$

och $D(X) = \sqrt{m}$.

6.6 Låt X vara likformigt fördelad på intervallet $[-0.05, 0.05]$, dvs $f_X(x) = 10$ då $-0.05 \le x \le 0.05$. Sätt $Y = 1/(0.5 + X)$. De möjliga värdena på Y ges av $[20/9,\ 20/11] = [1.818,\ 2.222]$. Väntevärdet är

$$
\begin{aligned}
E(Y) &= E\left(\frac{1}{0.5 + X}\right) = \int \frac{1}{0.5 + x}\, f_X(x)\, dx = \int_{-0.05}^{0.05} \frac{1}{0.5 + x}\, 10\, dx = 10\left[\ln(0.5 + x)\right]_{-0.05}^{0.05} \\
&= 10\ln(11/9) \approx 2.0067
\end{aligned}
$$

och variansen fås genom att först beräkna

$$
\begin{aligned}
E\left(Y^2\right) &= E\left(\left(\frac{1}{0.5+X}\right)^2\right) = \int \frac{1}{(0.5+x)^2}\, f_X(x)\, dx = \int_{-0.05}^{0.05} \frac{1}{(0.5+x)^2}\, 10\, dx \\
&= 10\left[\frac{-1}{0.5+x}\right]_{-0.05}^{0.05} = \frac{400}{99}
\end{aligned}
$$

så

$$V(Y) = E\left(Y^2\right) - (E(Y))^2 = \frac{400}{99} - (10\ln(11/9))^2 \approx 0.013531.$$

6.7 Den stokastiska variabeln X har tätheten

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2\sigma^2}.$$

Bestäm $E(|X|)$.

Alternativ 1: Nu har $|X|$ fördelningsfunktion

$$F_{|X|}(t) = P(|X| \le t) = P(-t \le X \le t) = P(-t < X \le t) = F_X(t) - F_X(-t).$$

Deriveras fördelningsfunktionen fås tätheten

$$f_{|X|}(t) = \frac{d}{dt}\left[F_X(t) - F_X(-t)\right] = f_X(t) - f_X(-t)(-1) = f_X(t) + f_X(-t).$$

för $t \ge 0$. Om X har täthet

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-t^2/2\sigma^2}.$$

så har $|X|$ täthet

$$f_{|X|}(t) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-t^2/2\sigma^2} + \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(-t)^2/2\sigma^2} = \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-t^2/2\sigma^2}.$$

Alltså är väntevärdet

$$
\begin{aligned}
E(|X|) &= \int x f_{|X|}(x)\, dx = \int_0^\infty x \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-x^2/2\sigma^2}\, dx = \{\text{Subst. } u = x^2/2\} \\
&= \int_0^\infty \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-u/\sigma^2}\, dx = \left[\frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-u/\sigma^2}(-\sigma^2)\right]_0^\infty = \sqrt{\frac{2}{\pi}}\sigma.
\end{aligned}
$$

Alternativ 2: Ur fördelningen för X får vi

$$
\begin{aligned}
E(|X|) &= \int |x|\, f_X(x)\, dx = \int_{-\infty}^\infty |x| \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-x^2/2\sigma^2}\, dx \\
&= \{\text{Jämn integrand, symmetriskt intervall}\} \\
&= 2\int_0^\infty |x| \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-x^2/2\sigma^2}\, dx = \int_0^\infty x \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-x^2/2\sigma^2}\, dx = \{\text{Subst. } u = x^2/2\} \\
&= \int_0^\infty \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-u/\sigma^2}\, dx = \left[\frac{\sqrt{2}}{\sqrt{\pi}\sigma} \mathrm{e}^{-u/\sigma^2}(-\sigma^2)\right]_0^\infty = \sqrt{\frac{2}{\pi}}\sigma.
\end{aligned}
$$

6.8 Ur fördelningsfunktionen $F_X(x) = x^2/\pi$ för $0 \le x \le \sqrt{\pi}$ bestäms

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}\frac{1}{\pi}x^2 = \frac{2x}{\pi}.$$

för $0 \le x \le \sqrt{\pi}$. Med $Y = \sin(X^2)$ så är

$$\begin{aligned} E(Y) &= E\left(\sin(X^2)\right) = \int \sin(x^2) f_X(x)\,dx = \int_0^{\sqrt{\pi}} \sin(x^2)\,\frac{2x}{\pi}\,dx = \{\text{Subst. } u = x^2\} \\ &= \int_0^{\pi} \sin(u)\,\frac{1}{\pi}\,du = \frac{1}{\pi}\left[-\cos(u)\right]_0^{\pi} = \frac{2}{\pi} \end{aligned}$$

och

$$\begin{aligned} E\left(Y^2\right) &= E\left(\sin^2(X^2)\right) = \int \sin^2(x^2) f_X(x)\,dx = \int_0^{\sqrt{\pi}} \sin^2(x^2)\,\frac{2x}{\pi}\,dx = \{\text{Subst. } u = x^2\} \\ &= \int_0^{\pi} \sin^2(u)\,\frac{1}{\pi}\,du = \int_0^{\pi} \frac{1-\cos(2u)}{2\pi}\,du = \left[\frac{u - \sin(2u)/2}{2\pi}\right]_0^{\pi} = \frac{1}{2}. \end{aligned}$$

Alltså är

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

och $D(X) = \sqrt{V(X)} = 0.30776$.

6.9 Låt Y_1 vara funktionstiden för seriekopplingen av komponent 1 och 2. Då är $Y_1 = \min(X_1, X_2)$ och

$$P(Y_1 > t) = P(\min(X_1, X_2) > t) = P(X_1 > t, X_2 > t) = P(X_1 > t)\,P(X_2 > t).$$

Då X_1 och X_2 är exponentialfördelade så är

$$P(X_1 > t) = \int_t^{\infty} f_{X_1}(x)\,dx = \int_t^{\infty} \frac{1}{a}\mathrm{e}^{-x/a}\,dx = \mathrm{e}^{-t/a},$$

$t \ge 0$, och $P(X_2 > t) = \mathrm{e}^{-t/a}$. Alltså är

$$P(Y_1 > t) = \mathrm{e}^{-t/a} \cdot \mathrm{e}^{-t/a} = \mathrm{e}^{-2t/a}$$

och

$$F_{Y_1}(t) = P(Y_1 \le t) = 1 - P(Y_1 > t) = 1 - \mathrm{e}^{-2t/a}$$

för $t \ge 0$. På samma sätt, med Y_2 som funktionstiden för seriekopplingen av komponent 3 och 4, är $F_{Y_2}(t) = 1 - \mathrm{e}^{-2t/a}$. Funktionstiden för systemet Y ges av

$$Y = \max(Y_1, Y_2)$$

och

$$\begin{aligned} F_Y(t) &= P(Y \le t) = P(\max(Y_1, Y_2) \le t) = P(Y_1 \le t, Y_2 \le t) \\ &= P(Y_1 \le t)\,P(Y_2 \le t) = (1 - \mathrm{e}^{-2t/a})^2 \end{aligned}$$

för $t \ge 0$. Derivering ger

$$f_Y(t) = \frac{d}{dt}F_Y(t) = 2(1 - \mathrm{e}^{-2t/a})\mathrm{e}^{-2t/a}\frac{2}{a} = \frac{4}{a}(\mathrm{e}^{-2t/a} - \mathrm{e}^{-4t/a}).$$

Nu har Y väntevärde

$$\begin{aligned} E(Y) &= \int t f_Y(t)\,dt = \int_0^{\infty} t\,\frac{4}{a}(\mathrm{e}^{-2t/a} - \mathrm{e}^{-4t/a})\,dt = \frac{4}{a}\left[t\,(\mathrm{e}^{-2t/a}(-a/2) - \mathrm{e}^{-4t/a}(-a/4))\right]_0^{\infty} \\ &\quad + 2\int_0^{\infty} \mathrm{e}^{-2t/a}\,dt - \int_0^{\infty} \mathrm{e}^{-4t/a}\,dt = 2 \cdot a/2 - a/4 = \tfrac{3}{4}a. \end{aligned}$$

Med $a = 1000$ är $E(Y) = 750$.

Notera att då Y alltid är positiv så

$$\begin{aligned}
E(Y) &= \int_0^\infty y f_Y(y)\,dy = \int_0^\infty \int_0^y f_Y(x)\,dxdy = \int_0^\infty \int_x^\infty f_Y(x)\,dydx = \int_0^\infty P(Y \geq x)\,dx \\
&= \int_0^\infty 1 - F_Y(x)dx.
\end{aligned}$$

Alltså

$$E(Y) = \int_0^\infty 1 - (1 - \mathrm{e}^{-2t/a})^2\,dt = \int_0^\infty 2\mathrm{e}^{-2t/a} - \mathrm{e}^{-4t/a}\,dt = \tfrac{3}{4}a$$

vilket man fick utan att bestämma $f_Y(t)$.

6.10 Med X som temperaturen i Celsius är temperaturen uttryckt i Farenheit

$$Y = \frac{9}{5}X + 32,$$

där X har väntevärde $E(X) = 17.0$ och standardavvikelse $D(X) = 1.6$. Via lineariteten för väntevärden får man

$$E(Y) = E\left(\frac{9}{5}X + 32\right) = \frac{9}{5}E(X) + 32 = 62.6°\mathrm{F}$$

och

$$V(Y) = V\left(\frac{9}{5}X + 32\right) = V\left(\frac{9}{5}X\right) = \frac{9^2}{5^2}V(X) = \frac{81}{25}(D(X))^2 = 8.2944$$

och $D(Y) = \sqrt{V(Y)} = 2.88°$F.

6.11 Med X som mängden tvätt och Y som priset för tvätten fås

$$Y = 5 + 10X,$$

där X har täthetsfunktion $f_X(x) = x/18$ då $0 \leq x \leq 6$. Här följer tre alternativ för att bestämma $E(Y)$ och $V(Y)$.

Alternativ 1: Via fördelningen för Y. Om de möjliga värdena på X är $S_X = [0, 6]$ så ges de möjliga värdena för $Y = 5 + 10X$ av $S_Y = [5, 65]$. För $t \in S_Y$ så är fördelningsfunktionen för Y

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

Deriveras denna fås tätheten

$$f_Y(t) = \frac{d}{dt}F_Y(t) = \frac{d}{dt}F_X((t-5)/10) = f_X((t-5)/10)\frac{1}{10} = \frac{(t-5)/10}{18} \cdot \frac{1}{10} = \frac{t-5}{1800}$$

för $5 \leq t \leq 65$. Med denna täthet har Y väntevärde

$$E(Y) = \int t f_Y(t)\,dt = \int_5^{65} t\frac{t-5}{1800}dt = \frac{1}{1800}\left[\frac{t^3}{3} - 5\frac{t^2}{2}\right]_5^{65} = 45$$

och varians

$$\begin{aligned}
V(Y) &= E\left((Y - E(Y))^2\right) = E\left((Y-45)^2\right) = \int (t-45)^2 f_Y(t)\,dt \\
&= \int_5^{65} (t-45)^2\frac{t-5}{1800}dt = \frac{1}{1800}\int_5^{65} t^3 - 95t^2 + 2475t - 10125dt \\
&= \frac{1}{1800}\left[\frac{t^4}{4} - 95\frac{t^3}{3} + 2475\frac{t^2}{2} - 10125t\right]_5^{65} = 200
\end{aligned}$$

Alternativ 2: Via fördelningen för X. Eftersom $E\left(g(X)\right) = \int g(x) f_X(x)\,dx$ så är

$$\begin{aligned} E\left(Y\right) &= E\left(5+10X\right) = \int (5+10x) f_X(x)\,dx = \int_0^6 (5+10x)\frac{x}{18}\,dx \\ &= \frac{1}{18}\left[5\frac{x^2}{2} + 10\frac{x^3}{3}\right]_0^6 = 45. \end{aligned}$$

och

$$\begin{aligned} V\left(Y\right) &= E\left((Y-E\left(Y\right))^2\right) = E\left(((5+10X)-45)^2\right) = E\left((10X-40)^2\right) \\ &= E\left(100X^2 - 800X + 1600\right) = \int (100x^2 - 800x + 1600) f_X(x)\,dx \\ &= \int_0^6 (100x^2 - 800x + 1600)\frac{x}{18}\,dx = \frac{1}{18}\left[100\frac{x^4}{4} - 800\frac{x^3}{3} + 1600\frac{x^2}{2}\right]_0^6 = 200 \end{aligned}$$

Alternativ 3: Via $E\left(X\right)$ och $V\left(X\right)$ och räknelagarna för väntevärden och varianser. För X gäller att

$$E\left(X\right) = \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \int_0^6 x \cdot \frac{x}{18} = \left[\frac{x^3}{3\cdot 18}\right]_0^6 = 4$$

och

$$\begin{aligned} V\left(X\right) &= E\left((X-E\left(X\right))^2\right) = \int_{-\infty}^{\infty} (x-4)^2 f_X(x)\,dx = \int_0^6 (x-4)^2\frac{x}{18}\,dx \\ &= \frac{1}{18}\left[\frac{x^4}{4} - \frac{8x^3}{3} + \frac{16x^2}{2}\right]_0^6 = 2. \end{aligned}$$

Alltså

$$E\left(Y\right) = E\left(5+10X\right) = 5 + 10E\left(X\right) = 5 + 40 = 45 \text{ kronor},$$
$$V\left(Y\right) = V\left(5+10X\right) = V\left(10X\right) = (10)^2 \cdot V\left(X\right) = 200$$

och $D\left(Y\right) = \sqrt{V\left(Y\right)} = \sqrt{200} = 10\sqrt{2}$.

6.12 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.

Nu är

$$E\left(X\right) = \sum_k kP\left(X=k\right) = (-1)P\left(X=-1\right) + 0P\left(X=0\right) + 1P\left(X=1\right) = 0.$$

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

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

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

6.13 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

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

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,

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

och

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

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

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

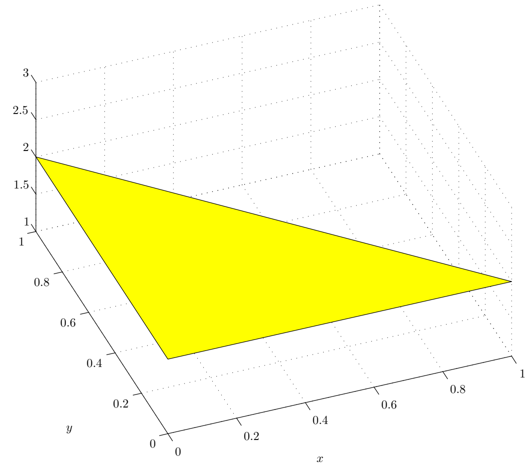
6.14 Vi börjar med att bestämma konstanten c: $f_{X,Y}(x,y)=c$.

$$1 = \iint_{x,y} f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^{1-y} c\,dx\,dy = \int_0^1 c(1-y)\,dy = c\left[y-\frac{y^2}{2}\right]_0^1 = \frac{c}{2},$$

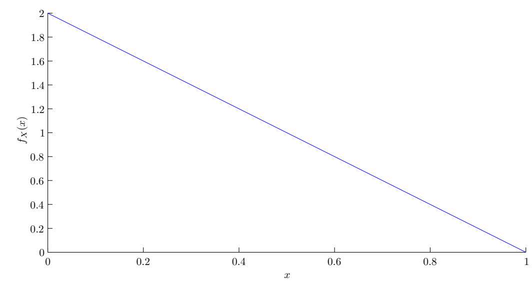
dvs $f_{X,Y}(x,y)=2$ på triangeln. Marginalfördelningen för X fås till

$$f_X(x) = \int f_{X,Y}(x,y)\,dy = \int_0^{1-x} 2\,dy = 2(1-x).$$

för $0\le x\le 1$. Av symmetriskäl är $f_Y(y)=2(1-y)$ för $0\le y\le 1$.



Tätheten $f_{X,Y}(x,y)$



Tätheten $f_X(x)$

Notera vidare att $2 = f_{X,Y}(x,y) \neq f_X(x) f_Y(y) = 2(1-x) \cdot 2(1-y)$ så vi har formellt visat att X och Y är beroende.

För X gäller att

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

och

$$V(X) = E\left((X - E(X))^2\right) = \int_{-\infty}^{\infty} \left(x - \frac{1}{3}\right)^2 f_X(x)\,dx = \int_0^1 \left(x - \frac{1}{3}\right)^2 2(1-x)dx = \frac{1}{18}.$$

Av symmetriskäl har även Y väntevärde 2/3 och varians 1/18. Vidare

$$\begin{aligned} E(XY) &= \iint_{x,y} xy f_{X,Y}(x,y)\,dx\,dy = \int_0^1 \int_0^{1-y} xy2\,dx\,dy = \int_0^1 2y\left[\frac{x^2}{2}\right]_0^{1-y} dy \\ &= \int_0^1 y(1-y)^2\,dy = \frac{1}{12} \end{aligned}$$

så

$$C(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{12} - \frac{1}{3} \cdot \frac{1}{3} = -\frac{1}{36}.$$

Slutligen, korrelationskoefficienten fås till

$$\rho_{X,Y} = \frac{C(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{-1/36}{\sqrt{\frac{1}{18} \cdot \frac{1}{18}}} = \frac{-1}{2}.$$

7.1 Lår 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

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

så

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

7.2 Låt X beskriva längden av en typisk planka. 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...

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

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

och

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

och

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

b) ...plankorna sågas till individuellt och får oberoende längder $X_1, \ldots, X_n$. 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

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

metod b) ger mindre standardavvikelse för den totala längden Y.

7.3 Låt X_1 och X_2 beskriva resultaten av två oberoende tärningskast. Då är $P(X_i = k) = 1/6$ för $k = 1, 2, \ldots, 6$ och

$$E(X_i) = \sum_{k=1}^{6} kP(X = k) = \frac{7}{2} = 3.5$$

och

$$V(X_i) = E\left(X^2\right) - (E(X))^2 = \sum_{k=1}^{6} k^2 P(X = k) - (3.5)^2 = \frac{35}{12}$$

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 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

$$\begin{aligned} V(X_1X_2) &= E\left((X_1X_2)^2\right) - (E(X_1x_2))^2 = E\left(X_1^2X_2^2\right) - \left(\frac{49}{4}\right)^2 = \{\text{oberoende}\} \\ &= E\left(X_1^2\right)E\left(X_2^2\right) - \frac{2401}{16}. \end{aligned}$$

Men $E\left(X_1^2\right) = V(X_1) + (E(X_1))^2 = \frac{35}{12} + \left(\frac{7}{2}\right)^2 = \frac{91}{6}$ så

$$V(Y) = V(X_1X_2) = \frac{91}{6} \cdot \frac{91}{6} - \frac{2401}{16} = \frac{11515}{144}.$$

Slutligen

$$C(X,Y) = E(XY) - E(X)E(Y) = E((X_1 + X_2)X_1X_2) - E(X_1 + X_2)E(X_1X_2)$$

där

$$\begin{aligned} E((X_1 + X_2)X_1X_2) &= E\left(X_1^2X_2 + X_2^2X_1\right) = \{\text{oberoende}\} \\ &= E\left(X_1^2\right)E(X_2) + E\left(X_2^2\right)E(X_1) = \frac{91}{6}\frac{7}{2} + \frac{91}{6}\frac{7}{2} = \frac{637}{6}, \end{aligned}$$

dvs

$$C(X,Y) = E((X_1 + X_2)X_1X_2) - E(X_1 + X_2)E(X_1X_2) = \frac{637}{6} - 7\frac{49}{4} = \frac{245}{12}.$$

Korrelationen fås till

$$\rho_{X,Y} = \frac{C(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{245/12}{\sqrt{\frac{35}{6}\frac{11515}{144}}} = \sqrt{\frac{42}{47}}.$$

7.4 Vi låter X_A och X_B vara två stokastiska variabler sådana att

$$E(X_A) = E(X_B) = m \quad \text{och} \quad V(X_A) = \sigma_A^2,\ V(X_B) = \sigma_B^2.$$

Vi låter $Y = aX_A + bX_B$. Då är

$$E(Y) = E(aX_A + bX_B) = aE(X_A) + bE(X_B) = am + bm = (a+b)m$$

vilket ger $E(Y) = m$ 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

$$\begin{aligned} V(Y) &= V((1-b)X_A + bX_B) = \{\text{oberoende}\} = (1-b)^2V(X_A) + b^2V(X_B) \\ &= (1-b)^2\sigma_A^2 + b^2\sigma_B^2. \end{aligned}$$

Denna minimeras med avseende på b genom lösning av

$$0 = \frac{d}{db}V(Y) = -2(1-b)\sigma_A^2 + 2b\sigma_B^2 = 2\left[(\sigma_A^2 + \sigma_B^2)b - \sigma_A^2\right]$$

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 andraderivatan ger att detta verkligen är ett minimum.

7.5 Med $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ så är variansen

$$\begin{aligned} V\left(\overline{X}\right) &= V\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^n V(X_i) = \frac{1}{n^2}\sum_{i=1}^n (D(X_i))^2 \\ &= \frac{1}{49}\left(4 \cdot (0.20)^2 + 2 \cdot (0.30)^2 + 1 \cdot (0.40)^2\right) = \frac{1}{98} \end{aligned}$$

så $D\left(\overline{X}\right) = 1/\sqrt{98} = 0.1010$. Notera att om bara A:s mätningar utnyttjas fås

$$V\left(\frac{1}{4}\sum_{i=1}^{4} X_i\right) = \{\text{oberoende}\} = \frac{1}{4^2}\sum_{i=1}^{4}(D\left(X_i\right))^2 = \frac{(0.2)^2}{4} = 0.01$$

och standardavvikelsen är $\sqrt{0.01} = 0.10$, något med bättre precision.

7.6 Låt X, Y och Z vara de oberoende normalfördelade stokastiska variabler som beskriver hastighetsvektorns komponenter, dvs de har väntevärde 0 och standardavvikelse $\sqrt{k_B T/M}$. Då är

$$E\left(\frac{1}{2}M|v|^2\right) = E\left(\frac{1}{2}M\left(X^2+Y^2+Z^2\right)\right) = \frac{1}{2}M\left(E\left(X^2\right)+E\left(Y^2\right)+E\left(Z^2\right)\right),$$

men

$$E\left(X^2\right) = V\left(X\right) + \left(E\left(X\right)\right)^2 = \frac{k_B T}{M} + 0^2 = \frac{k_B T}{M},$$

detsamma gäller $E\left(Y^2\right)$ och $E\left(Z^2\right)$, så

$$E\left(\frac{1}{2}M|v|^2\right) = \frac{1}{2}M\left(\frac{k_B T}{M} + \frac{k_B T}{M} + \frac{k_B T}{M}\right) = \frac{3}{2}k_B T.$$

Med hjälp av stora talens lag kan vi tolka resultatet som att medelvärdet av partiklarnas rörelseenergi är proportionell mot absoluta temperaturen.

7.7 Beteckna $E\left(X\right) = m_x$ och $E\left(Y\right) = m_y$. Upprepad användning av $V\left(Z\right) = E\left(Z^2\right) - \left(E\left(Z\right)\right)^2$ ger

$$\begin{aligned}
V\left(XY\right) &= E\left((XY)^2\right) - \left(E\left(XY\right)\right)^2 = E\left(X^2Y^2\right) - \left(E\left(XY\right)\right)^2 = \{\text{oberoende}\} \\
&= E\left(X^2\right)E\left(Y^2\right) - \left(E\left(X\right)E\left(Y\right)\right)^2 = E\left(X^2\right)E\left(Y^2\right) - m_x^2 m_y^2 \\
&= (V\left(X\right) + m_x^2)(V\left(Y\right) + m_y^2) - m_x^2 m_y^2 \\
&= V\left(X\right)V\left(Y\right) + V\left(X\right)m_y^2 + V\left(Y\right)m_x^2 + m_x^2 m_y^2 - m_x^2 m_y^2 \\
&= V\left(X\right)V\left(Y\right) + m_y^2 V\left(X\right) + m_x^2 V\left(Y\right).
\end{aligned}$$

7.8 Inför beteckningar $V\left(X\right) = \sigma_x^2 = 2$ och $V\left(Y\right) = \sigma_y^2 = 1$. Notera att eftersom

$$\rho_{X,Y} = \frac{C\left(X,Y\right)}{\sqrt{V\left(X\right)V\left(Y\right)}} = \frac{C\left(X,Y\right)}{\sigma_x \sigma_y}$$

så är

$$C\left(X,Y\right) = \sigma_x \sigma_y \rho_{X,Y} = \sqrt{2}\,\sqrt{1}\,\frac{1}{3} = \frac{\sqrt{2}}{3}.$$

Alternativ 1:

$$\begin{aligned}
V\left(2X-3Y\right) &= V\left(2X\right) + V\left(-3Y\right) + 2C\left(2X,-3Y\right) \\
&= 2^2 V\left(X\right) + (-3)^2 V\left(Y\right) + 2\cdot 2\cdot(-3)C\left(X,Y\right) \\
&= 4\sigma_x^2 + 9\sigma_y^2 - 12C\left(X,Y\right) = 17 - 4\sqrt{2} = 11.34.
\end{aligned}$$

Alternativ 2:

$$\begin{aligned}
V\left(2X-3Y\right) &= C\left(2X-3Y, 2X-3Y\right) \\
&= C\left(2X,2X\right) + C\left(2X,-3Y\right) + C\left(-3Y,2X\right) + C\left(-3Y,-3Y\right) \\
&= 2\cdot 2C\left(X,X\right) + 2\cdot(-3)C\left(X,Y\right) + (-3)\cdot 2C\left(Y,X\right) + (-3)\cdot(-3)C\left(Y,Y\right) \\
&= 4\sigma_x^2 - 6\frac{\sqrt{2}}{3} - 6\frac{\sqrt{2}}{3} + 9\sigma_y^2 = 17 - 4\sqrt{2} = 11.34.
\end{aligned}$$

7.9 Notera att för $i \neq j$ så är

$$\rho = \frac{C(X_i, X_j)}{\sqrt{V(X_i)\,V(X_j)}} = \frac{C(X_i, X_j)}{\sigma \cdot \sigma} \quad \text{så} \quad C(X_i, X_j) = \rho\sigma^2.$$

Vi vet att $V\left(\overline{X}\right) \geq 0$. Detta medför att

$$\begin{aligned}
0 &\leq V\left(\overline{X}\right) = C\left(\overline{X}, \overline{X}\right) = C\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2} C\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} C(X_i, X_j) = \frac{1}{n^2}\left(\sum_{i=1}^{n} V(X_i) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} C(X_i, X_j)\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} \sigma^2 + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \sigma^2\rho\right) = \frac{1}{n^2}\left(n\sigma^2 + n(n-1)\sigma^2\rho\right) = \frac{\sigma^2}{n}\left(1 + (n-1)\rho\right).
\end{aligned}$$

Vilket ger

$$0 \leq 1 + (n-1)\rho$$

eller $\rho \geq -1/(n-1)$.

7.10 Den genererande funktionen till en diskret stokastisk variabel definieras av $G(s) = E\left(s^X\right)$. Om två oberoende stokastiska variabler X_1 och X_2 har genererande funktioner $G_{X_1}(s)$ och $G_{X_2}(s)$ så har summan $X_1 + X_2$ genererande funktion

$$G_{X_1+X_2}(s) = E\left(s^{X_1+X_2}\right) = E\left(s^{X_1}\, s^{X_2}\right) = \{\text{oberoende}\} = E\left(s^{X_1}\right) \cdot E\left(s^{X_2}\right) = G_{X_1}(s) G_{X_2}(s).$$

En Poissonfördelad stokastisk variabel har $P(X = k) = \frac{m^k}{k!}\mathrm{e}^{-m}$ för $k = 0, 1, 2, \ldots$ där $m > 0$. Då har X genererande funktion

$$\begin{aligned}
G_X(s) &= E\left(s^X\right) = \sum_k s^k P(X = k) = \sum_{k=0}^{\infty} s^k \frac{m^k}{k!}\mathrm{e}^{-m} = \sum_{k=0}^{\infty} \frac{(sm)^k}{k!}\mathrm{e}^{-ms}\mathrm{e}^{ms-m} \\
&= \mathrm{e}^{m(s-1)} \sum_{k=0}^{\infty} \underbrace{\frac{(sm)^k}{k!}\mathrm{e}^{-ms}}_{P(Y=k)} = \mathrm{e}^{m(s-1)} \underbrace{\sum_{k=0}^{\infty} P(Y = k)}_{=1} = \mathrm{e}^{m(s-1)}.
\end{aligned}$$

där Y var Poissonfördelad med parameter sm.

Om X och Y är oberoende och Poissonfördelade med parametrar m_x och m_y har $X + Y$ genererande funktion

$$G_{X+Y}(s) = G_X(s) G_Y(s) = \mathrm{e}^{m_x(s-1)}\mathrm{e}^{m_y(s-1)} = \mathrm{e}^{(m_x+m_y)(s-1)} = \mathrm{e}^{m(s-1)}$$

där $m = m_x + m_y$, dvs den genererande funktionen för en Poissonfördelad stokastisk variabel.

7.11 Låt X vara likformigt fördelad på intervallet $[-0.05, 0.05]$, dvs $f_X(x) = 10$ då $-0.05 \leq x \leq 0.05$. Av symmetriskäl är $E(X) = 0$ och variansen är

$$V(X) = E\left((X-0)^2\right) = \int_{-0.05}^{0.05} x^2\, f_X(x)\, dx = 10\left[\frac{x^3}{3}\right]_{-0.05}^{0.05} = \frac{0.01}{12}.$$

Med $g(x) = 1/(0.5 + x)$ så är $Y = 1/(0.5 + X) = g(X)$ och

$$g'(x) = \frac{d}{dx} g(x) = \frac{-1}{(0.5+x)^2}.$$

Med Gauss approximationsformler fås

$$
\begin{aligned}
E(Y) &= E(g(X)) \approx g(E(X)) = g(0) = \frac{1}{0.5+0} = 2 \\
V(Y) &= V(g(X)) \approx [g'(E(X))]^2 V(X) = [g'(0)]^2 \frac{0.01}{12} = (-4)^2 \cdot \frac{0.01}{12} = \frac{0.04}{3}.
\end{aligned}
$$

7.12 Låt X beskriva uppmätt diffraktionsvinkel. Modell: $E(X) = \theta$ och $D(X) = \sigma$. Med $g(x) = c/\sin(x)$ så är $Y = c/\sin(X) = g(X)$. Nu är

$$
g'(x) = \frac{d}{dx} g(x) = \frac{-c}{\sin^2(x)} \cos(x).
$$

Med Gauss approximationsformler fås

$$
\begin{aligned}
E(Y) &= E(g(X)) \approx g(E(X)) = g(\theta) = \frac{c}{\sin(\theta)} \\
V(Y) &= V(g(X)) \approx [g'(E(X))]^2 V(X) = [g'(\theta)]^2 \sigma^2 = \left(\frac{-c}{\sin^2(\theta)} \cos(\theta)\right)^2 \cdot \sigma^2 = \frac{c^2 \cos^2(\theta)}{\sin^4(\theta)} \sigma^2
\end{aligned}
$$

7.13 Låt X beskriva en mätning av klotets diameter. Modell: $E(X) = d$ och $V(X) = \sigma^2$. Medelvärdet av n oberoende mätningar har väntevärde

$$
E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} d = d
$$

och varians

$$
V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}
$$

så $D\left(\overline{X}\right) = \sigma/\sqrt{n}$. Volymen av ett klot med diameter x ges av

$$
g(x) = \frac{\pi}{6} x^3
$$

som har derivata

$$
g'(x) = \frac{d}{dx} g(x) = \frac{\pi}{2} x^2.
$$

Med $Y = g(\overline{X})$ och Gauss approximationsformler fås

$$
\begin{aligned}
E(Y) &= E(g(X)) \approx g(E(X)) = g(d) = \frac{\pi}{6} d^3 \\
V(Y) &= V(g(X)) \approx [g'(E(X))]^2 V(X) = [g'(d)]^2 \frac{\sigma^2}{n} = \left(\frac{\pi}{2} d^2\right)^2 \frac{\sigma^2}{n} = \frac{\pi^2 d^4 \sigma^2}{4n}
\end{aligned}
$$

så $D(Y) \approx d^2 \pi\sigma/\sqrt{4n}$.

7.14 Beteckna $E(X) = m_x$, $E(Y) = m_y$, $V(X) = \sigma_x^2$ och $V(Y) = \sigma_y^2$ Upprepad användning av $V(Z) = E\left(Z^2\right) - (E(Z))^2$ ger

$$
\begin{aligned}
V(XY) &= E\left((XY)^2\right) - (E(XY))^2 = E\left(X^2 Y^2\right) - (E(XY))^2 = \{\text{oberoende}\} \\
&= E\left(X^2\right) E\left(Y^2\right) - (E(X)E(Y))^2 = E\left(X^2\right) E\left(Y^2\right) - m_x^2 m_y^2 \\
&= (\sigma_x^2 + m_x^2)(\sigma_y^2 + m_y^2) - m_x^2 m_y^2 \\
&= \sigma_x^2\sigma_y^2 + \sigma_x^2 m_y^2 + \sigma_y^2 m_x^2 + m_x^2 m_y^2 - m_x^2 m_y^2 \\
&= \sigma_x^2\sigma_y^2 + m_y^2\sigma_x^2 + m_x^2\sigma_y^2.
\end{aligned}
$$

Med funktionen $g(x,y) = xy$ så skall $V(g(X,Y))$ approximeras. Nu är

$$g_x(x,y) = \frac{\partial}{\partial x} g(x,y) = y \qquad g_y(x,y) = \frac{\partial}{\partial y} g(x,y) = x.$$

Med Gauss approximationsformler och oberoendet fås

$$\begin{aligned} V(XY) = V(g(X,Y)) &\approx [g_x(E(X), E(Y))]^2 V(X) + [g_y(E(X), E(Y))]^2 V(Y) \\ &= m_y^2 \sigma_x^2 + m_x^2 \sigma_y^2. \end{aligned}$$

7.15 Låt X beskriva ljusstrålens vinkel mot normalen där $E(X) = \pi/4$ och $V(X) = \sigma^2$. Träffpunktens position Y ges av $Y = d\tan(X)$. Med funktionen $g(x) = d\tan(x)$ får man med Gauss approximationsformler

$$E(Y) = E(g(X)) \approx g(E(X)) = d\tan(E(X)) = d\tan(\pi/4) = d.$$

Vidare,

$$g'(x) = \frac{d}{dx} g(x) = \frac{d}{dx} d\tan(x) = d(1 + \tan^2(x))$$

så

$$V(Y) = V(g(X)) \approx [g'(E(X))]^2 V(X) = [d(1 + \tan^2(E(X)))]^2 \sigma^2 = 4d^2\sigma^2,$$

dvs. $D(Y) \approx 2d\sigma = 2 \cdot 10 \cdot 0.05 = 1$.

7.16 Med funktionen

$$g(l,\tau) = \frac{4\pi^2 l}{\tau^2}$$

så är

$$g_l(l,\tau) = \frac{\partial}{\partial l} g(l,\tau) = \frac{4\pi^2}{\tau^2} \qquad g_\tau(l,\tau) = \frac{\partial}{\partial \tau} g(l,\tau) = \frac{-8\pi^2 l}{\tau^3}.$$

Låt L och T vara de oberoende stokastiska variabler som beskriver en mätning av pendellängden och svängningstiden. Inför beteckningar: $E(L) = l$, $E(T) = \tau$, $V(L) = \sigma_l^2$ och $V(T) = \sigma_\tau^2$.

Med Gauss approximationsformler och oberoendet fås

$$\begin{aligned} E\left(\frac{4\pi^2 L}{T^2}\right) = E(g(L,T)) &\approx g(E(L), E(T)) = g(l,\tau) = \frac{4\pi^2 l}{\tau^2} \\ V\left(\frac{4\pi^2 L}{T^2}\right) = V(g(L,T)) &\approx [g_l(E(L), E(T))]^2 V(L) + [g_\tau(E(L), E(T))]^2 V(T) \\ &= \left(\frac{4\pi^2}{\tau^2}\right)^2 \sigma_l^2 + \left(\frac{-8\pi^2 l}{\tau^3}\right)^2 \sigma_\tau^2. \end{aligned}$$

Med siffror $l = 1.0031$, $\tau = 2.0084$, $\sigma_l = 0.0002$ och $\sigma_\tau = 0.00043$ fås

$$\begin{aligned} E\left(\frac{4\pi^2 L}{T^2}\right) &\approx 9.8176 \\ V\left(\frac{4\pi^2 L}{T^2}\right) &\approx 2.1504 \cdot 10^{-5} \\ D\left(\frac{4\pi^2 L}{T^2}\right) &\approx 0.00464. \end{aligned}$$

7.17 Funktionen

$$g(V,B,r) = \frac{2V}{B^2 r^2}$$

har derivatorna

$$\begin{aligned} g_V(V,B,r) &= \frac{\partial}{\partial V} g(V,B,r) = \frac{2}{B^2r^2} \\ g_B(V,B,r) &= \frac{\partial}{\partial B} g(V,B,r) = \frac{-4V}{B^3r^2} \\ g_r(V,B,r) &= \frac{\partial}{\partial r} g(V,B,r) = \frac{-4V}{B^2r^3}. \end{aligned}$$

Med $\tilde{V}$, $\tilde{B}$ och $\tilde{r}$ som oberoende stokastiska variabler som beskriver uppmätt spänningsdifferens, flödestäthet och radie respektive.

Med modellen

$$V = E\left(\tilde{V}\right) = 299 \qquad B = E\left(\tilde{B}\right) = 1.173 \cdot 10^{-3} \qquad r = E\left(\tilde{r}\right) = 4.9 \cdot 10^{-2}$$
$$\sigma_V = D\left(\tilde{V}\right) = 1.2 \qquad \sigma_B = D\left(\tilde{B}\right) = 5.0 \cdot 10^{-7} \qquad \sigma_r = D\left(\tilde{r}\right) = 2.2 \cdot 10^{-5}$$

får man med Gauss approximationsformler att

$$\begin{aligned} E\left(g(\tilde{V},\tilde{B},\tilde{r})\right) &\approx g(E\left(\tilde{V}\right), E\left(\tilde{B}\right), E\left(\tilde{r}\right)) = g(V,B,r) = \frac{2V}{B^2r^2} = 1.8101 \cdot 10^{11} \\ V\left(g(\tilde{V},\tilde{B},\tilde{r})\right) &\approx \left[g_V(V,B,r)\right]^2 \sigma_V^2 + \left[g_B(V,B,r)\right]^2 \sigma_B^2 + \left[g_r(V,B,r)\right]^2 \sigma_r^2 \\ &= \left(\frac{2}{B^2r^2}\right)^2 \sigma_V^2 + \left(\frac{-4V}{B^3r^2}\right)^2 \sigma_B^2 + \left(\frac{-4V}{B^2r^3}\right)^2 \sigma_r^2 = 5.7801 \cdot 10^{17} \\ D\left(g(\tilde{V},\tilde{B},\tilde{r})\right) &\approx 7.6027 \cdot 10^8 \end{aligned}$$

7.18 Funktionen

$$g(h,R,r) = \pi h(R^2 - r^2)$$

har derivatorna

$$\begin{aligned} g_h(h,R,r) &= \frac{\partial}{\partial h} g(h,R,r) = \pi(R^2 - r^2) \\ g_R(h,R,r) &= \frac{\partial}{\partial R} g(h,R,r) = 2\pi hR \\ g_r(h,R,r) &= \frac{\partial}{\partial r} g(h,R,r) = -2\pi hr. \end{aligned}$$

Låt $\tilde{R}$, $\tilde{r}$ och $\tilde{h}$ vara de stokastiska variabler som beskriver en mätning av yttre radie, inre radie och höjd av en hålcylinder. Modell:

$$h = E\left(\tilde{h}\right) = 5.0 \qquad R = E\left(\tilde{R}\right) = 8.2 \qquad r = E\left(\tilde{r}\right) = 3.0$$
$$\sigma_h = D\left(\tilde{h}\right) = 0.1 \qquad \sigma_R = D\left(\tilde{R}\right) = 0.2 \qquad \sigma_r = D\left(\tilde{r}\right) = 0.2$$
$$C\left(\tilde{h},\tilde{R}\right) = 0 \qquad C\left(\tilde{R},\tilde{r}\right) = 0.01 \qquad C\left(\tilde{r},\tilde{h}\right) = 0$$

Volymen V av hålcylindern ges av $V = g(\tilde{h},\tilde{R},\tilde{r})$. Med Gauss approximationsformler fås

$$\begin{aligned} E\left(V\right) = E\left(g(\tilde{h},\tilde{R},\tilde{r})\right) &\approx g(E\left(\tilde{h}\right), E\left(\tilde{R}\right), E\left(\tilde{r}\right)) = \pi h(R^2 - r^2) = 914.83 \\ V\left(V\right) = V\left(g(\tilde{h},\tilde{R},\tilde{r})\right) &\approx g_h\, g_h C\left(\tilde{h},\tilde{h}\right) + g_h\, g_R C\left(\tilde{h},\tilde{R}\right) + g_h\, g_r C\left(\tilde{h},\tilde{r}\right) \\ &+ g_R\, g_h C\left(\tilde{R},\tilde{h}\right) + g_R\, g_R C\left(\tilde{R},\tilde{R}\right) + g_R\, g_r C\left(\tilde{R},\tilde{r}\right) \\ &+ g_r\, g_h C\left(\tilde{r},\tilde{h}\right) + g_r\, g_R C\left(\tilde{r},\tilde{R}\right) + g_r\, g_r C\left(\tilde{r},\tilde{r}\right) \\ &= g_h^2\sigma_h^2 + g_R^2\sigma_R^2 + g_r^2\sigma_r^2 + 2g_R\, g_r C\left(\tilde{R},\tilde{r}\right) = 2859 \\ D\left(V\right) &\approx \sqrt{2859} = 53.47. \end{aligned}$$

8.1 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 $a = \lambda_{0.025} = 1.9600$.

8.2 Låt X vara N(m,σ) med $m = 5$ och $\sigma = 2$. Då är $Z = (X - m)/\sigma$ N$(0,1)$ och

a)
$$P(X \leq 6) = P\left(\frac{X - m}{\sigma} \leq \frac{6 - m}{\sigma}\right) = P\left(Z \leq \frac{1}{2}\right) = \Phi(0.5) = 0.6915.$$

b)
$$\begin{aligned} P(1.8 < X < 7.2) &= P(1.8 < X \leq 7.2) = P\left(\frac{1.8 - m}{\sigma} < \frac{X - m}{\sigma} \leq \frac{7.2 - m}{\sigma}\right) \\ &= P(-1.6 < Z \leq 1.1) = \Phi(1.1) - \Phi(-1.6) = \Phi(1.1) - (1 - \Phi(1.6)) \\ &= 0.80953. \end{aligned}$$

c) a sådan att $P(X \leq a) = 0.05$ bestäms ur

$$\begin{aligned} 1 - \Phi(\lambda_{0.05}) = 0.05 &= P(X \leq a) = P\left(\frac{X - m}{\sigma} \leq \frac{a - m}{\sigma}\right) = P\left(Z \leq \frac{a - m}{\sigma}\right) \\ &= \Phi\left(\frac{a - m}{\sigma}\right) = 1 - \Phi\left(-\frac{a - m}{\sigma}\right). \end{aligned}$$

Alltså är

$$-\frac{a - m}{\sigma} = \lambda_{0.05}$$

eller

$$a = m - \lambda_{0.05}\sigma = 5 - 1.6449 \cdot 2 = 1.7103.$$

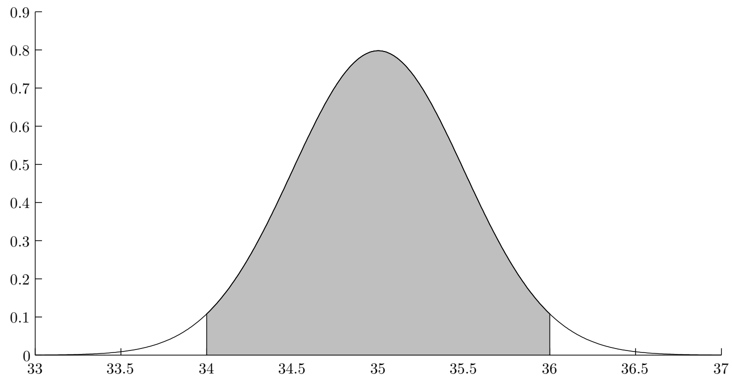
8.3 Låt X beskriva mängden kaffe i säcken, X är N(m,σ) där $m = 35$ och $\sigma = 0.5$.

a) Då är sannolikheten att säcken räcker till minst 36 burkar

$$\begin{aligned} P(X > 36) &= P\left(\frac{X - m}{\sigma} > \frac{36 - m}{\sigma}\right) = 1 - \Phi\left(\frac{36 - m}{\sigma}\right) \\ &= 1 - \Phi(2) = 1 - 0.97725 = 0.02275. \end{aligned}$$

b) Då är sannolikheten att säcken räcker till 34 men inte till 36 burkar

$$\begin{aligned} P(34 < X < 36) &= P\left(\frac{34 - m}{\sigma} < \frac{X - m}{\sigma} < \frac{36 - m}{\sigma}\right) = \Phi(2) - \Phi(-2) \\ &= \Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1 = 2 \cdot 0.97725 - 1 = 0.9545. \end{aligned}$$



Tätheten $f_X(x)$ för N$(35, 0.5)$.

8.4 Låt X beskriva diametern för en tillverkad axel. Modell: X är N(m, σ). Vi bestämmer m och σ ur ekvationerna

$$\begin{aligned} 0.08 &= P(X > 1.01) = P\left(\frac{X-m}{\sigma} > \frac{1.01-m}{\sigma}\right) = 1 - \Phi\left(\frac{1.01-m}{\sigma}\right) \\ 0.02 &= P(X < 0.99) = P\left(\frac{X-m}{\sigma} < \frac{0.99-m}{\sigma}\right) = \Phi\left(\frac{0.99-m}{\sigma}\right) = 1 - \Phi\left(-\frac{0.99-m}{\sigma}\right). \end{aligned}$$

Alltså är

$$\begin{cases} \dfrac{1.01-m}{\sigma} = \lambda_{0.08} \\ -\dfrac{0.99-m}{\sigma} = \lambda_{0.02} \end{cases}$$

och

$$\begin{aligned} m &= \frac{1.01\lambda_{0.02} + 0.99\lambda_{0.08}}{\lambda_{0.02} + \lambda_{0.08}} = 1.0019 \\ \sigma &= \frac{1.01 - 0.99}{\lambda_{0.02} + \lambda_{0.08}} = 0.0057823. \end{aligned}$$

Notera att kvantilerna $\lambda_{0.08} = 1.4051$ och $\lambda_{0.02} = 2.0537$ normalt inte är tillgängliga i tabellform utan fås via dator eller invertering av tabell för $\Phi(\cdot)$.

8.5 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

$$m = E(Y) = E(X_1 + X_2) = E(X_1) + E(X_2) = 200 + 100 = 300$$

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(m, σ) = N$(300, 25)$. Sökt är

$$P(Y > 310) = P\left(\frac{Y-m}{\sigma} > \frac{310-m}{\sigma}\right) = 1 - \Phi(0.4) = 0.34458.$$

8.6 Låt X beskriva den inre diametern hos en hylsa och Y den yttre diametern hos en tapp. Modell: X är N$(10, 0.2)$ och Y är N$(9.5, 0.1)$ där X och Y är oberoende stokastiska variabler. För att en utvald hylsa och tapp skall passa varandra så skall $X - Y > 0.2$ och $X - Y < 1$. Med $W = X - Y$ så är W normalfördelad med väntevärde

$$m = E(W) = E(X - Y) = E(X) - E(Y) = 10 - 9.5 = 0.5$$

och varians

$$\sigma^2 = V(W) = V(X - Y) = \{\text{oberoende}\} = V(X) + (-1)^2 V(Y) = 0.2^2 + 0.1^2 = 0.05$$

dvs W är N$(0.5, \sqrt{0.05})$. Vidare

$$\begin{aligned} P(0.2 < X - Y < 1) &= P(0.2 < W < 1) = P\left(\frac{0.2 - m}{\sigma} < \frac{W - m}{\sigma} < \frac{1 - m}{\sigma}\right) \\ &= \Phi\left(\frac{1 - m}{\sigma}\right) - \Phi\left(\frac{0.2 - m}{\sigma}\right) = \Phi\left(\sqrt{5}\right) - \Phi\left(-3/\sqrt{5}\right) \\ &= \Phi\left(\sqrt{5}\right) - 1 + \Phi\left(3/\sqrt{5}\right) = 0.8975. \end{aligned}$$

8.7 Låt X beskriva tiden för relä 1 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 $Y - X < 0$. Med $W = Y - X$ så är W normalfördelad med väntevärde

$$m = E(W) = E(Y - X) = E(Y) - E(X) = 1.5 - 1 = 0.5$$

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$(0.5, \sqrt{0.05})$. Vidare

$$\begin{aligned} P(Y - X < 0) &= P(W < 0) = P\left(\frac{W - m}{\sigma} < \frac{0 - m}{\sigma}\right) = \Phi(-m/\sigma) = \Phi\left(-\sqrt{5}\right) \\ &= 1 - \Phi\left(\sqrt{5}\right) = 1 - 0.9873 = 0.0127. \end{aligned}$$

8.8 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 $X_1, \ldots, X_n$ är oberoende och fördelade som X. Då är Y normalfördelad med väntevärde

$$m = E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = nE(X) = 10 \cdot 70 = 700$$

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$(m, \sigma) =$ N$(700, \sqrt{1000})$. Sökt är

$$P(Y > 800) = P\left(\frac{Y - m}{\sigma} > \frac{800 - m}{\sigma}\right) = 1 - \Phi\left(\sqrt{10}\right) = 0.0007827.$$

8.9 Låt X beskriva vikten av en typisk pappersrulle. Modell: X är N(m, σ). Låt Y beskriva vikten för $n = 10$ pappersrullar,

$$Y = X_1 + X_2 + \cdots + X_n,$$

där $X_1, \ldots, X_n$ är oberoende och fördelade som X. Då är Y normalfördelad med väntevärde

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = nE(X) = 10m = 1000$$

och varians

$$V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{n} V(X_i) = nV(X) = 10\sigma^2 = 20^2$$

dvs $m = 100$ och $\sigma = \sqrt{40}$. Vi vill bestämma x så att $P(X > x) = 0.90$. Detta görs genom

$$\begin{aligned} 1 - \Phi(\lambda_{0.10}) = 0.10 &= P(X \le x) = P\left(\frac{X-m}{\sigma} \le \frac{x-m}{\sigma}\right) = \Phi\left(\frac{x-m}{\sigma}\right) \\ &= 1 - \Phi\left(-\frac{x-m}{\sigma}\right). \end{aligned}$$

Alltså är

$$-\frac{x-m}{\sigma} = \lambda_{0.10}$$

eller

$$x = m - \lambda_{0.10}\sigma = 100 - 1.2816 \cdot \sqrt{40} = 91.895.$$

8.10 Låt X_1 och X_2 beskriva de två avståndsmätningarna. Modell: X_1 och X_2 är oberoende och X_i är $\mathrm{N}(a_i, \sigma_i)$. Med $Y = 0.80X_1 + 0.60X_2$ så är Y normalfördelad med väntevärde

$$b = E(Y) = E(0.80X_1 + 0.60X_2) = 0.80E(X_1) + 0.60E(X_2) = 0.8a_1 + 0.6a_2$$

och varians

$$\begin{aligned} \sigma^2 &= V(Y) = V(0.80X_1 + 0.60X_2) = \{\text{oberoende}\} = (0.80)^2 V(X_1) + (0.60)^2 V(X_2) \\ &= 0.64\sigma_1^2 + 0.36\sigma_2^2. \end{aligned}$$

Med $a_1 = 4000$ och $a_2 = 3000$ fås $b = 5000$, $\sigma_1 = 3 \cdot 10^{-6} a_1 = 0.012$, $\sigma_2 = 3 \cdot 10^{-6} a_2 = 0.009$ och

$$\sigma^2 = 0.64\sigma_1^2 + 0.36\sigma_2^2 = 0.00012132$$

dvs $\sigma = 0.011015$. Sökt är

$$\begin{aligned} P(|Y - b| \le 0.02) &= P(-0.02 \le Y - b \le 0.02) = P\left(-\frac{0.02}{\sigma} \le \frac{Y-b}{\sigma} \le \frac{0.02}{\sigma}\right) \\ &= \Phi\left(\frac{0.02}{\sigma}\right) - \Phi\left(-\frac{0.02}{\sigma}\right) = 2\Phi\left(\frac{0.02}{\sigma}\right) - 1 = 2\Phi(1.8158) - 1 \\ &= 0.9306. \end{aligned}$$

8.11 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,

$$Y = X_1 + X_2 + \cdots + X_n,$$

där $X_1, \ldots, X_n$ är oberoende och fördelade som X. Då har Y väntevärde

$$m = E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = nE(X) = 25 \cdot 10 = 250$$

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 CGS är Y approximativt $\mathrm{N}(m, \sigma) = \mathrm{N}(250, 2.5)$. Sökt är

$$P(Y > 255) = P\left(\frac{Y-m}{\sigma} > \frac{255-m}{\sigma}\right) \approx 1 - \Phi(2) = 0.02275.$$

8.12 (Något generaliserad.) Låt $U_1, U_2, \ldots, U_n$ vara n stycken oberoende $\mathrm{R}(0,1)$ fördelade stokastiska variabler. Ur formelsamling så är

$$E(U_i) = \frac{0+1}{2} = \frac{1}{2} \qquad V(U_i) = \frac{(1-0)^2}{12} = \frac{1}{12},$$

för $i = 1, \ldots, n$. Med

$$Y = m + \sigma\sqrt{\frac{12}{n}}\left(\sum_{i=1}^{n} U_i - \frac{n}{2}\right)$$

så är

$$\begin{aligned} E(Y) &= E\left(m + \sigma\sqrt{\frac{12}{n}}\left(\sum_{i=1}^{n} U_i - \frac{n}{2}\right)\right) = m + \sigma\sqrt{\frac{12}{n}}\left(\sum_{i=1}^{n} E(U_i) - \frac{n}{2}\right) \\ &= m + \sigma\sqrt{\frac{12}{n}}\left(n\frac{1}{2} - \frac{n}{2}\right) = m \end{aligned}$$

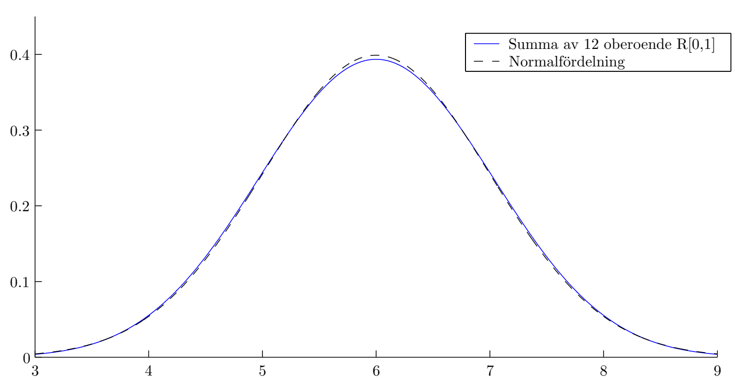
och, utnyttjandes oberoendet,

$$\begin{aligned} V(Y) &= V\left(m + \sigma\sqrt{\frac{12}{n}}\left(\sum_{i=1}^{n} U_i - \frac{n}{2}\right)\right) = V\left(\sigma\sqrt{\frac{12}{n}}\left(\sum_{i=1}^{n} U_i - \frac{n}{2}\right)\right) \\ &= \sigma^2\frac{12}{n}V\left(\sum_{i=1}^{n} U_i - \frac{n}{2}\right) = \sigma^2\frac{12}{n}V\left(\sum_{i=1}^{n} U_i\right) = \sigma^2\frac{12}{n}\sum_{i=1}^{n} V(U_i) = \sigma^2\frac{12}{n}\cdot\frac{n}{12} = \sigma^2. \end{aligned}$$

Enligt CGS så är $\sum_{i=1}^{n} U_i$ approximativ normalfördelad vilket innebär att den linjära transformationen

$$Y = m + \sigma\sqrt{\frac{12}{n}}\left(\sum_{i=1}^{n} U_i - \frac{n}{2}\right)$$

av $\sum_{i=1}^{n} U_i$ är approximativt normalfördelad. Med $n = 12$ fås uppgiftens påstående.



Fördelningen för $\sum_{i=1}^{12} U_i$ jämfört med en normalfördelning med samma väntevärde och varians som summan.

8.13 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

$$E(X) = \sum_k kP(X = k) = 0 \cdot 0.40 + \cdots 3 \cdot 0.10 = 1.1$$

och

$$V(X) = E\left((X - E(X))^2\right) = \sum_k (k - 1.1)^2 P(X = k) = 1.09.$$

Låt Y beskriva antalet barn i förskoleåldern i ett område med $n = 1000$ hushåll,

$$Y = X_1 + X_2 + \cdots + X_n,$$

där $X_1, \ldots, X_n$ är oberoende och fördelade som X. Då har Y väntevärde

$$m = E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = nE(X) = 1000 \cdot 1.1 = 1100$$

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 $\mathrm{N}(m, \sigma) = \mathrm{N}(1100, \sqrt{1090})$. Vi söker y så att $P(Y < y) = 0.90$.

$$1 - \Phi(\lambda_{0.10}) = 0.10 \quad = \quad P(Y > y) = P\left(\frac{Y - m}{\sigma} > \frac{y - m}{\sigma}\right) \approx 1 - \Phi\left(\frac{y - m}{\sigma}\right).$$

Alltså är approximativt

$$\frac{y - m}{\sigma} = \lambda_{0.10}$$

eller

$$y = m + \lambda_{0.10}\sigma = 1100 + 1.2816 \cdot \sqrt{1090} = 1142.3.$$

Alltså, bygg 1143 platser.

8.14 Låt $X_1, \ldots, X_n$ vara oberoende och likformigt fördelade på $[-0.5, 0.5]$. Då är

$$E(X_i) = 0 \quad \text{och} \quad V(X_i) = \frac{1}{2} = \sigma^2.$$

Med $Y = \sum_{i=1}^{n} X_i$ så har Y väntevärde

$$m = E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = nE(X) = 1000 \cdot 0 = 0$$

och varians

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{n} V(X_i) = nV(X) = n\sigma^2 = n/12.$$

Enligt CGS är Y för stora n approximativt $\mathrm{N}(m, \sigma) = \mathrm{N}(0, \sqrt{n/12})$. Alltså är

$$\begin{aligned} P\left(|Y| < K\sqrt{n}\right) &= P\left(-K\sqrt{n} < Y < K\sqrt{n}\right) \\ &= P\left(\frac{-K\sqrt{n}}{\sqrt{n/12}} < \frac{Y - 0}{\sqrt{n/12}} < \frac{K\sqrt{n}}{\sqrt{n/12}}\right) \approx 2\Phi\left(\sqrt{12/K}\right) - 1. \end{aligned}$$

Det sista högerledet är oberoende av n så felet $|Y - 0|$ växer i storleksordningen $\sqrt{n}$. Notera att n påverkar dock precisionen hos normalapproximationen.

8.15 Låt $X_1, X_2, \ldots, X_n$, $n = 50$, beskriva livslängderna för elektronrören. Modell: X_i är oberoende och exponentialfördelade med väntevärde $m = 200$. Den totala tiden som lagret räcker beskrivs av $T = \sum_{i=1}^{n} X_i$ som har väntevärde

$$E(T) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = nm = 50 \cdot 200 = 10000$$

och varians

$$V(T) = V\left(\sum_{i=1}^{n} X_i\right) = \{\text{ober.}\} = \sum_{i=1}^{n} \underbrace{V(X_i)}_{=m^2} = nm^2 = 50 \cdot 200^2 = 2 \cdot 10^6.$$

Vi säker tiden t sådan att $P(T > t) = 0.90$ (tänk!). Enligt CGS så är T approximativt normalfördelat så

$$0.90 = P(T > t) = P\left(\frac{T - 10000}{\sqrt{2 \cdot 10^6}} > \frac{t - 10000}{\sqrt{2 \cdot 10^6}}\right) \approx 1 - \Phi\left(\frac{t - 10000}{\sqrt{2 \cdot 10^6}}\right),$$

dvs $(t - 10000)/\sqrt{2 \cdot 10^6} \approx \lambda_{0.90} = -\lambda_{0.10} = -1.2816$, eller

$$t \approx 10000 - \lambda_{0.10}\sqrt{2 \cdot 10^6} = 10000 - 1.2816 \cdot 1414.2 = 8187.6$$

8.16 Låt $T_1, T_2, \ldots, T_n$ beskriva tiderna mellan kundernas ankomster där T_i är tiden mellan kund i och $i + 1$, för $i = 1, \ldots, n - 1$. Modell: T_i är oberoende och exponentialfördelade med parameter $m = 1.5$, dvs

$$E(T_i) = m = 1.5 \qquad V(T_i) = m^2 = 2.25.$$

Tiden X_n då kund n kommer in på kontoret beskrivs av

$$X_n = \sum_{i=1}^{n-1} T_i$$

som har väntevärde

$$E(X_n) = E\left(\sum_{i=1}^{n-1} T_i\right) = \sum_{i=1}^{n-1} E(T_i) = (n-1)m = 49 \cdot 1.5 = 73.5$$

och varians

$$V(X_n) = V\left(\sum_{i=1}^{n-1} T_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{n-1} V(T_i) = (n-1)m^2 = 49m^2,$$

dvs $D(X_n) = \sqrt{49m^2} = 7m = 10.5$. Enligt CGS är X_n approximativt $\mathrm{N}(73.5, 10.5)$. Alltså är

$$P(X_n < 65) = P\left(\frac{X_n - 73.5}{10.5} < \frac{65 - 73.5}{10.5}\right) \approx \Phi\left(\frac{65 - 73.5}{10.5}\right) = \Phi(-0.8095) = 0.20911.$$

8.17 Låt X beskriva en typisk förflyttning i x-led. Modell: en förflyttning sker till någon av de sex punkterna $(1,0,0)$, $(-1,0,0)$, $(0,1,0)$, $(0,-1,0)$, $(0,0,1)$, $(0,0,-1)$ med lika stor sannolikhet. Då har X sannolikhetsfunktion

$$p_X(-1) = \frac{1}{6} \qquad p_X(0) = \frac{4}{6} \qquad p_X(1) = \frac{1}{6}.$$

Då är

$$E(X) = \sum_k kP(X = k) = -1 \cdot \frac{1}{6} + 0 \cdot \frac{4}{6} + 1 \cdot \frac{1}{6} = 0$$

och

$$V(X) = E\left((X - E(X))^2\right) = \sum_k (k-0)^2 P(X = k) = \frac{1}{3}.$$

Låt Y beskriva den totala förflyttningen i x-led efter $n = 300$ steg,

$$Y = X_1 + X_2 + \cdots + X_n,$$

där $X_1, \ldots, X_n$ är oberoende och fördelade som X. Då har Y väntevärde

$$m = E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X) = 300 \cdot 0 = 0$$

och varians

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(X_i) = nV(X) = 300 \cdot \frac{1}{3} = 100.$$

dvs $D(Y) = 10$. Enligt CGS är Y approximativt $\mathrm{N}(m, \sigma) = \mathrm{N}(0, 10)$. Vi söker

$$\begin{aligned}
P(|Y| > 20) &= 1 - P(|Y| \le 20) = 1 - P(-20 \le Y \le 20) \\
&= 1 - P\left(\frac{-20-m}{\sigma} \le \frac{Y-m}{\sigma} \le \frac{20-m}{\sigma}\right) \\
&\approx 1 - \left(\Phi\left(\frac{20-m}{\sigma}\right) - \Phi\left(\frac{-20-m}{\sigma}\right)\right) \\
&= 1 - \Phi(2) + \Phi(-2) = 2(1 - \Phi(2)) = 0.0455.
\end{aligned}$$

9.1 Sannolikhetsfunktionen för U_i, $i = 1, \ldots, n$, är

$$p_{U_i}(x) = P(U_i = x) = \begin{cases} p & \text{för } x = 1 \\ 1-p & \text{för } x = 0 \\ 0 & \text{annars} \end{cases}$$

så

$$E(U_i) = \sum_k k \cdot P(U_i = k) = 0 \cdot P(U_i = 0) + 1 \cdot P(U_i = 1) = p.$$

Vidare,

$$E\left(U_i^2\right) = \sum_k k^2 \cdot P(U_i = k) = 0^2 \cdot P(U_i = 0) + 1^2 \cdot P(U_i = 1) = p$$

så

$$V(U_i) = E\left(U_i^2\right) - (E(U_i))^2 = p - p^2 = p(1-p).$$

Med X som det totala antalet gånger som A inträffat så är

$$X = \sum_{i=1}^n U_i$$

med

$$E(X) = E\left(\sum_{i=1}^n U_i\right) = \sum_{i=1}^n E(U_i) = np$$

och

$$V(X) = V\left(\sum_{i=1}^n U_i\right) = \{\text{oberoende}\} = \sum_{i=1}^n V(U_i) = np(1-p).$$

X är binomialfördelad.

9.2 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 Bin(n, p), dvs

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

för $k = 0, 1, 2, \ldots, n$. 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

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^{3} \binom{n}{k} p^k (1-p)^{n-k} = 1 - 0.94444 = 0.055556.$$

9.3 Låt X_1 beskriva antalet defekta varvräknare i urvalet från maskin 1 och X_2 motsvarande för maskin 2. Modell: X_1 och X_2 är oberoende och X_i är 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 om inte $p_1 = p_2$.

9.4 Låt X beskriva antalet fungerande komponenter bland de $n = 15$ komponenterna. Modell: komponenter fungerar oberoende av varandra med sannolikhet p så X är Bin(n, p), dvs

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

för $k = 0, 1, 2, \ldots, n$. Då är

$$P(\text{Syst. fungerar}) = P(X \geq m) = \sum_{k=m}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

9.5 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 $p \geq 1/3$.

9.6 Låt X beskriva antalet träffar av en måltavla i $n = 5$ försök. Antag att träffar sker oberoende av varandra och med sannolikhet $p = 1/3$ i varje försök. Då är X Bin(n,p) och

$$\begin{aligned}
P(2 \leq X \leq 3) &= P(X=2)+P(X=3) = \binom{n}{2}p^2(1-p)^{n-2}+\binom{n}{3}p^3(1-p)^{n-3} \\
&= \binom{5}{2}(1/3)^2(2/3)^3+\binom{5}{3}(1/3)^3(2/3)^2 = 10\cdot\frac{8}{3^5}+10\cdot\frac{4}{3^5} = \frac{40}{81} = 0.4938
\end{aligned}$$

9.7 (Jämför med uppgift 2.17). Av $N = 100$ distinkta enheter är $s = 6$ defekta. Av $n = 5$ på måfå utvalda enheter låt X vara antalet defekta om urvalet skedde utan återläggning. Då är X hypergeometriskt fördelad och

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

för $k = 0,1,\ldots,5$. Med $p = s/N = 6/100$ som andelen defekta enheter i urnan så är

$$E(X) = np = 5\cdot 0.06 = 0.30 \qquad V(X) = np(1-p)\frac{N-n}{N-1} = 5\cdot 0.06\cdot 0.94\,\frac{100-5}{100-1} = 0.27061$$

dvs $D(X) = \sqrt{V(X)} = 0.5202$. Vidare så är

$$P(X \leq 1) = P(X=0)+P(X=1) = \frac{54891018+18297006}{75287520} = \frac{435643}{448140} = 0.97211.$$

9.8 Ur en befolkning av storlek N tas n personer på måfå. Med p som andelen rökare i befolkningen låt X [Y] beskriva antalet rökare i urvalet om det skedde med [utan] återläggning. Modell: X är Bin(n,p) och Y är Hypergeometriskt fördelad (N,n,p). Vi vet att

$$E(X) = np \qquad V(X) = np(1-p)$$

och

$$E(Y) = np \qquad V(Y) = \underbrace{np(1-p)}_{=V(X)}\frac{N-n}{N-1} = V(X)\,\frac{N-n}{N-1}$$

dvs $V(Y) < V(X)$ om $n > 1$. Med $p^* = X/n$ och $\hat{p} = Y/n$ så är

$$E(p^*) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = p \quad \text{och} \quad E(\hat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = p$$

så båda skattningarna är väntevärdesriktiga. Vidare,

$$V(p^*) = E\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{p(1-p)}{n}$$

och

$$V(\hat{p}) = V\left(\frac{Y}{n}\right) = \frac{1}{n^2}V(Y) = \frac{p(1-p)}{n}\,\frac{N-n}{N-1} = V(p^*)\,\frac{N-n}{N-1}$$

så $V(\hat{p}) < V(p^*)$ om $n > 1$. Skattningen $\hat{p}$ är effektivare än p^*.

9.9 Låt X beskriva antalet kunder som vill köpa en högtalare under en dag, där X är Poissonfördelad med väntevärde $m = 2$ kunder/dag. Då är

$$\begin{aligned} P(X > 4) &= 1 - P(X \le 4) = 1 - \sum_{k=0}^{4} P(X = k) = 1 - \sum_{k=0}^{4} \frac{m^k}{k!}\mathrm{e}^{-m} \\ &= 1 - \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \cdots + \frac{2^4}{4!}\right)\mathrm{e}^{-2} = 1 - 7\mathrm{e}^{-2} = 0.052653 \end{aligned}$$

9.10 Låt X beskriva antalet anrop till en växel under ett tidsintervall. Modell: Antag att det finns n abonnenter. Dessa ringer oberoende av varandra och med sannolikhet p under tidsintervallet. Då är X $\mathrm{Bin}(n,p)$ och

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

för $k = 0, 1, 2, \ldots, n$. Låt $n \to \infty$ och $p \to 0$ så att $E(X) = np = m$ är konstant. Då är

$$\begin{aligned} \lim_{\substack{n\to\infty \\ p\to 0}} P(X = k) &= \lim_{\substack{n\to\infty \\ p\to 0}} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n\to\infty} \binom{n}{k} \left(\frac{m}{n}\right)^k \left(1 - \frac{m}{n}\right)^{n-k} \\ &= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} m^k n^{-k} \left(1 - \frac{m}{n}\right)^{n-k} \\ &= \lim_{n\to\infty} \underbrace{\frac{n!}{n^k(n-k)!}}_{\to 1} \frac{1}{k!}\, m^k \underbrace{\left(1 - \frac{m}{n}\right)^n}_{\to \mathrm{e}^{-m}} \underbrace{\left(1 - \frac{m}{n}\right)^{-k}}_{\to 1} = \frac{m^k}{k!}\mathrm{e}^{-m} \end{aligned}$$

för $k = 0, 1, 2, 3, \ldots$. Detta är en giltig sannolikhetsfördelning ty

$$\sum_{k=0}^{\infty} \frac{m^k}{k!}\mathrm{e}^{-m} = \mathrm{e}^{-m} \underbrace{\sum_{k=0}^{\infty} \frac{m^k}{k!}}_{=\mathrm{e}^m} = \mathrm{e}^{-m}\,\mathrm{e}^{m} = \mathrm{e}^0 = 1.$$

9.11 Låt $X_1, \ldots, X_4$ beskriva antalet samtal till telefon 1,...,4 under tidsintervallet. Modell: X_i är oberoende och Poissonfördelade med väntevärden

$$E(X_1) = E(X_2) = E(X_3) = 1 \qquad E(X_4) = 0.5.$$

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

$$\begin{aligned} P(Y \ge 3) &= 1 - P(Y < 3) = 1 - P(Y \le 2) = 1 - \sum_{k=0}^{2} P(Y = k) = 1 - \sum_{k=0}^{2} \frac{m^k}{k!}\mathrm{e}^{-m} \\ &= 1 - 0.32085 = 0.67915. \end{aligned}$$

9.12 Fel uppstår med intensitet $\lambda_A = 0.3$ fel/år för maskin A och med intensitet $\lambda_B = 0.1$ fel/år för maskin B. Låt X och Y beskriva antalet fel på maskin A resp B under en period om t år. Modell: X och Y är oberoende Poissonfördelade stokastiska variabler med väntevärden $\lambda_A t$ resp. $\lambda_B t$. För 4 maskiner av typ A och 10 av typ B låt W vara det totala antalet fel,

$$W = \sum_{i=1}^{4} X_i + \sum_{i=1}^{10} Y_i,$$

där variabler är oberoende och X_i är fördelade som X och Y_i som Y. Då är W en Poissonfördelad stokastisk variabel med väntevärde

$$m = E(W) = E\left(\sum_{i=1}^{4} X_i + \sum_{i=1}^{10} Y_i\right) = \sum_{i=1}^{4} E(X_i) + \sum_{i=1}^{10} E(Y_i) = (4\lambda_A + 10\lambda_B)t.$$

Med $t = 2$ år fås $m = 4.4$ och

$$P(W \geq 6) = 1 - P(W < 6) = 1 - \sum_{k=0}^{5} P(W = k) = 1 - \sum_{k=0}^{5} \frac{m^k}{k!} \mathrm{e}^{-m} = 1 - 0.71991 = 0.28009.$$

9.13 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 $\mathrm{Bin}(n, p)$. 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

$$E(X) = np = 100 = m \qquad V(X) = np(1-p) = 90 = \sigma^2$$

Då $V(X) \geq 10$ är approximation av binomialfördelningen med normalfördelningen tillåten, dvs X är approximativt $\mathrm{N}(m, \sigma) = \mathrm{N}(100, \sqrt{90})$. Då är

$$\begin{aligned} P(80 \leq X \leq 120) &= P\left(\frac{80-m}{\sigma} \leq \frac{X-m}{\sigma} \leq \frac{120-m}{\sigma}\right) \approx \Phi\left(\frac{120-m}{\sigma}\right) - \Phi\left(\frac{80-m}{\sigma}\right) \\ &= 2\Phi\left(20/\sqrt{90}\right) - 1 = 0.96499. \end{aligned}$$

Med halvkorrektion får man svaret

$$\begin{aligned} P(80 \leq X \leq 120) &= P(79.5 \leq X \leq 120.5) = P\left(\frac{79.5-m}{\sigma} \leq \frac{X-m}{\sigma} \leq \frac{120.5-m}{\sigma}\right) \\ &\approx \Phi\left(\frac{120.5-m}{\sigma}\right) - \Phi\left(\frac{79.5-m}{\sigma}\right) = 2\Phi\left(20.5/\sqrt{90}\right) - 1 = 0.96930. \end{aligned}$$

9.14 Låt X beskriva antalet sålda bakelser bland $n = 100$. Modell: med pris c kr/bakelse är sannolikheten för att en bakelse blir såld $p = \mathrm{e}^{-c/2}$ och bakelser blir sålda oberoende av varandra. Då är X binomialfördelad $\mathrm{Bin}(n, p) = \mathrm{Bin}(n, \mathrm{e}^{-c/2})$.

Då är det förväntade antalet sålda bakelser $E(X) = np$ och den förväntade inkomsten

$$E(cX) = cE(X) = cnp = cn\mathrm{e}^{-c/2}.$$

Lösning av

$$0 = \frac{d}{dc} E(cX) = \frac{d}{dc}\left[cn\mathrm{e}^{-c/2}\right] = n\mathrm{e}^{-c/2}\left(1 - \frac{c}{2}\right)$$

ger $c = 2$ kronor/bakelse.

Med $c = 2$ är $p = \mathrm{e}^{-c/2} = \mathrm{e}^{-1} = 0.36788$ och

$$P(X \geq 30) = \sum_{k=30}^{100} \binom{n}{k} p^k (1-p)^{n-k} = 0.93642.$$

Då $V(X) \geq 10$ är approximation av binomialfördelningen med normalfördelningen tillåten, dvs X är approximativt $\mathrm{N}(m, \sigma)$ där

$$m = E(X) = np = 36.788 \qquad \sigma^2 = V(X) = np(1-p) = 23.254.$$

Således,

$$\begin{aligned} P(X \geq 30) &= P\left(\frac{X-m}{\sigma} \geq \frac{30-m}{\sigma}\right) \approx 1 - \Phi\left(\frac{30-m}{\sigma}\right) = 1 - \Phi(-1.4076) \\ &= \Phi(1.4076) = 0.92038. \end{aligned}$$

Med halvkorrektion fås svaret:

$$\begin{aligned} P(X \geq 29.5) &= P\left(\frac{X-m}{\sigma} \geq \frac{29.5-m}{\sigma}\right) \approx 1 - \Phi\left(\frac{29.5-m}{\sigma}\right) = 1 - \Phi(-1.51133) \\ &= \Phi(1.51133) = 0.93464. \end{aligned}$$

9.15 Låt X beskriva antalet hopp åt höger bland $n = 100$ steg. Modell: X är $\mathrm{Bin}(n, p)$ med $p = 1/2$. Vi vet att

$$E(X) = np = 50 = m \qquad V(X) = np(1-p) = 25 = \sigma^2.$$

Då $V(X) \geq 10$ så är X approximativt normalfördelad, dvs approximativt $\mathrm{N}(50, 5)$.

Med X steg åt höger (och $n - X$ steg åt vänster) ges positionen efter n steg av

$$Y = X - (n - X) = 2X - n.$$

Vi söker

$$\begin{aligned} P(|Y| = 20) &= P(Y = -20) + P(Y = 20) = P(2X - n = -20) + P(2X - n = 20) \\ &= P(X = 40) + P(X = 60) = \binom{n}{40} p^{40}(1-p)^{n-40} + \binom{n}{60} p^{60}(1-p)^{n-60} \\ &= 0.021687. \end{aligned}$$

Med normalapproximation fås samma sannolikhet till

$$\begin{aligned} P(X = 40) + P(X = 60) &= P(39.5 \leq X \leq 40.5) + P(59.5 \leq X \leq 60.5) \\ &= P\left(\frac{39.5-m}{\sigma} \leq \frac{X-m}{\sigma} \leq \frac{40.5-m}{\sigma}\right) \\ &\quad + P\left(\frac{59.5-m}{\sigma} \leq \frac{X-m}{\sigma} \leq \frac{60.5-m}{\sigma}\right) \\ &\approx \Phi(-1.9) - \Phi(-2.1) + \Phi(-1.9) - \Phi(-2.1) \\ &= 2(\Phi(2.1) - \Phi(1.9)) = 0.021704. \end{aligned}$$

På samma sätt

$$\begin{aligned}
P(|Y| \geq 20) &= P(Y \leq -20) + P(Y \geq 20) = P(2X - n \leq -20) + P(2X - n \geq 20) \\
&= P(X \leq 40) + P(X \geq 60) = 1 - \sum_{k=41}^{59} \binom{n}{k} p^k (1-p)^{n-k} \\
&= 0.056888.
\end{aligned}$$

Med normalapproximation fås samma sannolikhet till

$$\begin{aligned}
P(X \leq 40) + P(X \geq 60) &= P(X \leq 40.5) + P(X \geq 59.5) \\
&= P\left(\frac{X-m}{\sigma} \leq \frac{40.5-m}{\sigma}\right) \\
&\quad + P\left(\frac{59.5-m}{\sigma} \leq \frac{X-m}{\sigma}\right) \\
&\approx \Phi(-1.9) + 1 - \Phi(1.9) \\
&= 2(1 - \Phi(1.9)) = 0.057433.
\end{aligned}$$

9.16 Låt X beskriva antalet bland de $n = 500$ tillfrågade som svarade ja. Modell: X är $\mathrm{Bin}(n, p)$ med $p = 0.52$ andelen ja-sägare i populationen. Vi söker $P\left(X < \frac{n}{2}\right)$. Vi vet att $E(X) = np = 260 = m$ och $V(X) = np(1-p)$. Om $V(X) \geq 10$ så är normalapproximation tillåten, som i detta fall:

$$V(X) = np(1-p) = 124.8 = \sigma^2.$$

Alltså är X approximativt $\mathrm{N}(m, \sigma)$ och

$$\begin{aligned}
P(X < n/2) &= P(X < 250) = P\left(\frac{X-m}{\sigma} < \frac{250-m}{\sigma}\right) \approx \Phi\left(\frac{250-m}{\sigma}\right) \\
&= \Phi(-0.89514) = 0.18536
\end{aligned}$$

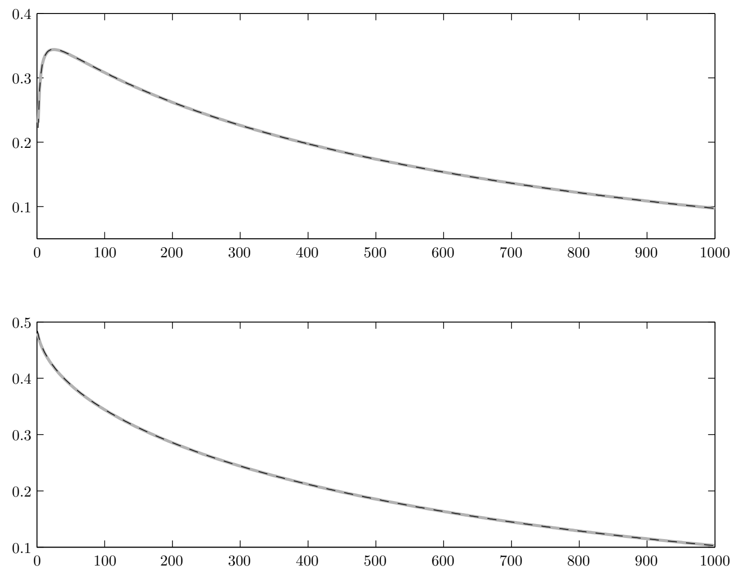
eller med halvkorrektion

$$P(X < 249.5) = P\left(\frac{X-m}{\sigma} < \frac{249.5-m}{\sigma}\right) \approx \Phi(-0.9399) = 0.17363.$$

Notera att kravet för normalapproximationen

$$V(X) = np(1-p) = n \cdot 0.52 \cdot 0.48 \geq 10$$

ger att n måste vara åtminstone 41. För olika värden på n bestäms $P(X < n/2)$ i figuren nedan.



Sannolikheten att av n utvalda det finns en minoritet ja-sägare när andelen ja-sägare i populationen är $p = 0.52$. Överst jämna n, nederst udda n. Heldragen linje är exakta uträkningar med binomialfördelningen streckad linje med normalapproximation och halvkorrektion.

9.17 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 $0 \le k \le s$ och $0 \le n - k \le N - s$. Eftersom $n/N < 0.10$ så kan fördelningen för X approximeras med binomialfördelningen, dvs

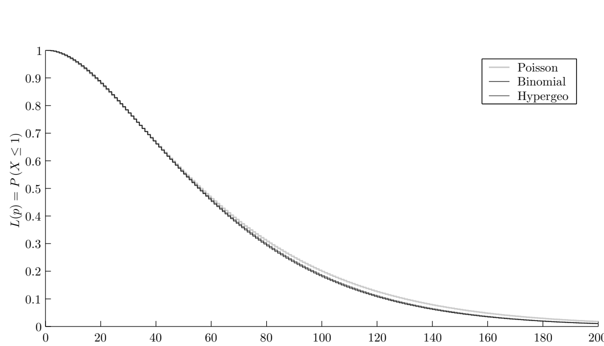
$$P(X = k) \approx \binom{n}{k} p^k (1-p)^{n-k}$$

för $k = 0, \ldots, n$, och då $p < 0.10$ kan binomialfördelningen approximeras med Poissonfördelningen med väntevärde np så

$$P(X = k) \approx \frac{(np)^k}{k!} \mathrm{e}^{-np}$$

för $k = 0, 1, \ldots,$. Nu är

$$L(p) = P(X \le 1) = P(X = 0) + P(X = 1) \approx \frac{(np)^0}{0!}\mathrm{e}^{-np} + \frac{(np)^1}{1!}\mathrm{e}^{-np} = (1 + np)\mathrm{e}^{-np}.$$



9.18 Låt $X_1, \ldots, X_n$ beskriva antalet larm under dagar $1, \ldots, n$. 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

$$m = E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = nE(X_i) = 365 \cdot \frac{1}{2} = 182.5.$$

Eftersom $m \geq 15$ så kan fördelningen för Y approximeras med en normalfördelning. Då är

$$\begin{aligned} P(Y > 200) &= P(Y > 200.5) = P\left(\frac{Y-m}{\sqrt{m}} > \frac{200.5-m}{\sqrt{m}}\right) \approx 1 - \Phi\left(\frac{200.5-m}{\sqrt{m}}\right) \\ &= 1 - \Phi(1.3324) = 0.091361. \end{aligned}$$

Exakt uträkning med Poissonfördelning ger $P(Y > 200) = 0.092831$.

9.19 Låt X beskriva antalet registrerade partiklar. Modell: X är Poissonfördelad med parameter $\lambda = ct$. Då är $E(X) = V(X) = \lambda$ och $D(X) = \sqrt{\lambda}$. Variationskoefficienten blir

$$\mathrm{R}(X) = \frac{D(X)}{E(X)} = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}}.$$

Det relativa slumpfelet

$$Y = \frac{X - E(X)}{E(X)} = \frac{1}{\lambda}X - 1$$

har väntevärde

$$E(Y) = \frac{1}{\lambda}E(X) - 1 = 0$$

och varians

$$V(Y) = \frac{1}{\lambda^2}V(X) = \frac{1}{\lambda}.$$

Om $E(X) = \lambda = 10^4 \geq 15$ så är normalapproximation av Poissonfördelning tillåten. Alltså är X approximativt normalfördelad och så även Y, dvs Y är approximativt $\mathrm{N}(0, 0.01)$ och

$$\begin{aligned} P(|Y| > 0.01) &= 1 - P(|Y| \leq 0.01) = 1 - P(-0.01 \leq Y \leq 0.01) \\ &= 1 - P\left(\frac{-0.01-0}{0.01} \leq \frac{Y-0}{0.01} \leq \frac{0.01-0}{0.01}\right) \approx 1 - (\Phi(1) - \Phi(-1)) \\ &= 2(1 - \Phi(1)) = 0.31731. \end{aligned}$$

Bestäm nu λ så att

$$P(|Y| \geq 0.02) \leq 0.05.$$

Detta ger att

$$0.05 = P(|Y| > 0.02) \approx 1 - (\Phi\left(0.02\sqrt{\lambda}\right) - \Phi\left(-0.02\sqrt{\lambda}\right)) = 2(1 - \Phi\left(0.02\sqrt{\lambda}\right))$$

eller

$$1 - \Phi(\lambda_{0.025}) = 0.025 = 1 - \Phi\left(0.02\sqrt{\lambda}\right),$$

dvs $\lambda = (50\lambda_{0.025})^2 = (50 \cdot 1.96)^2 = 9603.6$. Med $\lambda = ct = 10^3\, t$ fås $t = 9.603$.

9.20 I en 3000 meter lång tunnel är hastighetsbegränsningen 60 km/h = 1km/minut, dvs det tar 3 minuter att åka genom tunneln.

Antalet bilar X i tunneln vid en tidpunkt t är antalet bilar som kommit till tunneln under tidsintervallet $[t-3,t]$ av längd 3, där X är

$$X = X_1 + X_2,$$

en summa av två oberoende Poissonfördelade stokastiska variabler med väntevärde $10 \cdot 3 = 30$ och $8 \cdot 3 = 24$. Då är X Poissonfördelad med väntevärde $30 + 24 = 54$ och normalapproximation av X är tillåten. Vi får då

$$P(X \geq 60) = P\left(\frac{X - E(X)}{D(X)} \geq \frac{60 - E(X)}{D(X)}\right) \approx 1 - \Phi\left(\sqrt{2/3}\right) = 0.20711$$

eller med halvkorrektion

$$P(X \geq 59.5) = P\left(\frac{X - E(X)}{D(X)} \geq \frac{59.5 - E(X)}{D(X)}\right) \approx 1 - \Phi\left(\sqrt{121/216}\right) = 0.22709.$$

Exakt uträkning från Poissonfördelningen ger svaret 0.22404.

9.21 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 $\text{Bin}(n,p)$. Försäkringsbolagest vinst ges av

$$Y = n \cdot 10^4 - X10^6 = 10^6(1 - X).$$

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 $m = E(X) = np = 0.8$. Då är

$$1 - (P(X = 0) + P(X = 1)) \approx 1 - \left(\frac{m^0}{0!}\mathrm{e}^{-m} + \frac{m^1}{1!}\mathrm{e}^{-m}\right) = 1 - 0.80879 = 0.19121.$$

9.22 Låt X beskriva antalet felaktiga keramiska plattor i ett parti om n stycken. Modell: plattor är felaktiga med sannolikhet $p = 0.001$ och oberoende av varandra. X är $\text{Bin}(n,p)$. Om man lägger till k korrekta plattor till partiet beskriver

$$Y = k + (n - X)$$

antalet korrekta plattor. Vi söker k så att $P(Y < n) = 0.0001$ dvs $P(X > k) = 0.0001$. Med $n = 1000$ är $E(X) = np = 1$ och X är approximativt $\text{Po}(1)$ eftersom $p < 0.1$.

$$P(X > k) = 1 - P(X \leq k) = 1 - \sum_{i=0}^{k} \frac{1^i}{i!}\mathrm{e}^{-1} = 1 - \mathrm{e}^{-1} \sum_{i=0}^{k} \frac{1}{i!}$$

k	$\sum_{i=0}^{k} \frac{1}{i!}$	$P(X > k)$
0	1	0.63212
1	2	0.26424
2	2.5	0.080301
3	2.6667	0.018988
4	2.7083	0.0036598
5	2.7167	0.00059418
6	2.7181	$8.3241 \cdot 10^{-5}$

vilket ger $k = 6$. Med $n = 50000$ så är $m = E(X) = np = 50$ och $\sigma^2 = V(X) = np(1-p) = 49.95$ så normalapproximation är tillåten. Vi vill nu finna k så att

$$P(X > k) = P\left(\frac{X-m}{\sigma} > \frac{k-m}{\sigma}\right) = 1 - \Phi\left(\frac{k-m}{\sigma}\right) = 0.0001 = 1 - \Phi(\lambda_{0.0001})$$

dvs

$$\frac{k-m}{\sigma} = \lambda_{0.0001}$$

eller

$$k = m + \lambda_{0.0001}\sigma = 50 + 3.719 \cdot \sqrt{49.95} = 76.284$$

dvs 77 plattor. Med halvkorrektion fås:

$$P(X > k + 0.5) = P\left(\frac{X-m}{\sigma} > \frac{k-m+0.5}{\sigma}\right) = 1 - \Phi\left(\frac{k-m+0.5}{\sigma}\right) = 0.0001 = 1 - \Phi(\lambda_{0.0001})$$

dvs

$$k = m - 0.5 + \lambda_{0.0001}\sigma = 75.784$$

eller 76 plattor. Exakt uträkning ur binomialfördelningen ger svaren $k = 6$ (då $n = 1000$) och $k = 78$ (då $n = 50000$).

Med $n = 1000$ så är den förväntade reklamationskostnaden

$$E(5X) = 5E(X) = 5np = 5$$

vilket är mindre än kostnaden för att skicka med 6 plattor extra $= 6 \cdot 3 = 18$ kronor.

Med $n = 50000$ så är den förväntade reklamationskostnaden

$$E(5X) = 5E(X) = 5np = 250$$

vilket är större än kostnaden för att skicka med 77 plattor extra $= 77 \cdot 3 = 231$ kronor.

11.1 Under $n = 10$ dagar mättes avkastningen för den kemiska processen och man erhöll observationerna $x_1, \ldots, x_n$:

2.3 7 8 1 6 4 3.0 6 3.1 2.2

vilka ordnade i storleksordning ger $x_{(1)}, \ldots, x_{(n)}$:

2.1 2.2 2.3 2.4 2.6 2.6 2.7 2.8 3.0 3.1

där $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Spann för datamaterialet: $x_{(n)} - x_{(1)} = 1.0$. Om n är udda så är medianen: $x_{(n+1)/2}$, för jämnt n är medianen inte väldefinierad. Här: $x_{(5)} = x_{(6)} = 2.6$ så medianen$= 2.6$.

Aritmetiskt medelvärde:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2.58.$$

Stickprovsvarians:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\overline{x}^2\right) = 0.11067$$

Stickprovsstandardavvikelse:

$$s = \sqrt{s^2} = \sqrt{0.11067} = 0.33267.$$

11.2 Låt $x_1, \ldots, x_n$, $n = 10$, vara den första mätserien och $y_1, \ldots, y_m$, $m = 5$, den andra mätserien. Sammanslaget till en mätserie om $m+n$ värden betecknar vi observationerna enligt

$$\begin{array}{c} x_1\,,\, x_2\,,\,\cdots,\, x_n\,,\quad y_1\quad,\quad y_2\quad,\cdots,\quad y_m \\ \Downarrow \\ w_1,\, w_2,\,\cdots,\, w_n,\, w_{n+1},\, w_{n+2},\,\cdots,\, w_{n+m} \end{array}$$

Då blir

$$\begin{aligned} \overline{w} &= \frac{1}{n+m}\sum_{i=1}^{n+m} w_i = \frac{1}{n+m}\left(\sum_{i=1}^{n} w_i + \sum_{i=n+1}^{n+m} w_i\right) = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} y_i\right) \\ &= \frac{1}{n+m}\left(n\overline{x} + m\overline{y}\right) = \frac{n\overline{x}+m\overline{y}}{n+m} = 5310.6. \end{aligned}$$

Stickprovsvariansen för första stickprovet är

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$$

så $\sum_{i=1}^{n} x_i^2 = (n-1)s_x^2 + n\overline{x}^2 = 2.8228 \cdot 10^8$. På samma sätt är $\sum_{i=1}^{m} y_i^2 = (m-1)s_y^2 + m\overline{y}^2 = 422782341$ så

$$\sum_{i=1}^{n+m} w_i^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} y_i^2 = 7.0506 \cdot 10^8.$$

Alltså,

$$s_w^2 = \frac{1}{n+m-1}\left(\sum_{i=1}^{n+m} w_i^2 - (n+m)\overline{w}^2\right) = 19.39$$

och $s_w = \sqrt{s_w^2} = 4.4$.

11.3 Vi har $n = 800$ observationer med

$$\overline{x} = 10.161 \quad \text{och} \quad s_x = 32.356.$$

Låt x_1 vara den felstämplade observationen, dvs $x_1 = 924$ istället för 9.24. Med $y_1, \ldots, y_n$ som korrekt observationsserie är

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\left(y_1 - x_1 + \sum_{i=1}^{n} x_i\right) = \frac{y_1 - x_1}{n} + \overline{x} = \frac{9.24 - 924}{800} + 10.161 = 9.0175.$$

Nu är

$$s_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$$

så

$$\sum_{i=1}^{n} x_i^2 = (n-1)s_x^2 + n\overline{x}^2 = (800-1)1046.9 + 800 \cdot (10.161)^2 = 919078.414864$$

Alltså är

$$\begin{aligned} s_y^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right) = \frac{1}{n-1}\left(y_1^2 - x_1^2 + \sum_{i=1}^{n} x_i^2 - n\overline{y}^2\right) \\ &= \frac{1}{799}\left(9.24^2 - 924^2 + 919078.414864 - 800 \cdot 9.2462^2\right) = 0.41906 \end{aligned}$$

och $s_y = 0.64735$.

11.4 Dataserien $x_1, \ldots, x_n$, $n = 30$:

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

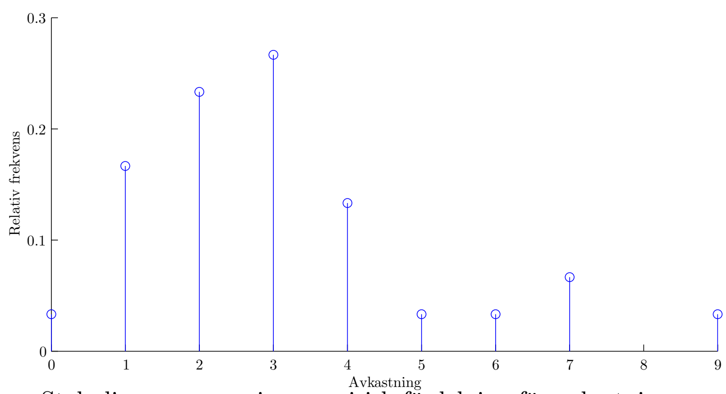
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 3.1$$

och stickprovsstandardavvikelse

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 2.0401.$$

Frekvenstabell:

Värde	Absolut frekvens	Relativ frekvens	Ackumulerad 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



Stolpdiagram som visar empirisk fördelning för avkastningen baserad på $n = 30$ observationer.

11.5 Med observationerna $x_1, \ldots, x_n$ grupperade i frekvenstabellen

Antal fläckar:	0	1	2	3	4	5
Frekvens:	22	18	7	2	0	1

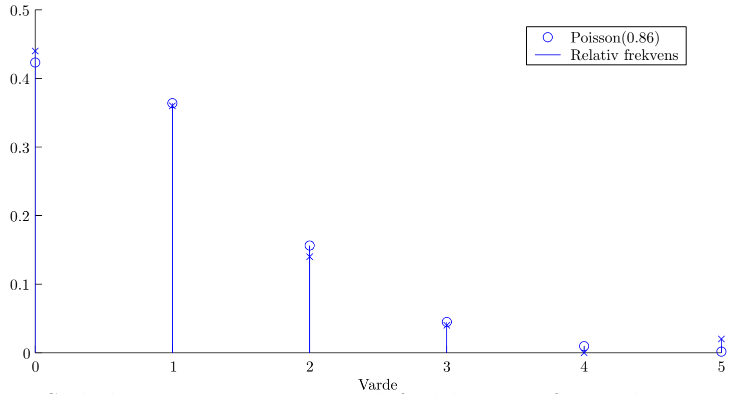
kan medelvärdet räknas ut

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(0 \cdot 22 + 1 \cdot 18 + \cdots + 5 \cdot 1) = 0.86$$

och stickprovsvariansen

$$s_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) = \frac{1}{n-1}\big((0^2 \cdot 22 + 1^2 \cdot 18 + \cdots + 5^2 \cdot 1) - n\overline{x}^2\big) = 1.0404$$

ger $s_x = 1.02$.



Stolpdiagram som visar empirisk fördelning jämfört med Poissonfördelningen med parameter $m = 0.86$.

11.6 Datamängden $x_1, \ldots, x_n$, $n = 50$,

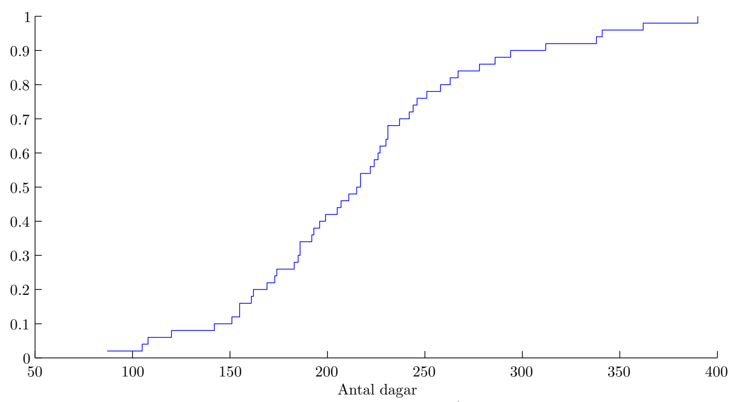
87	105	108	120	142	151	155	155	161	162
169	173	174	183	185	186	186	192	193	196
199	205	207	211	215	217	217	222	224	226
227	230	231	231	237	242	244	246	251	258
263	267	278	286	294	312	338	341	362	390

har medelvärde

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 217.08$$

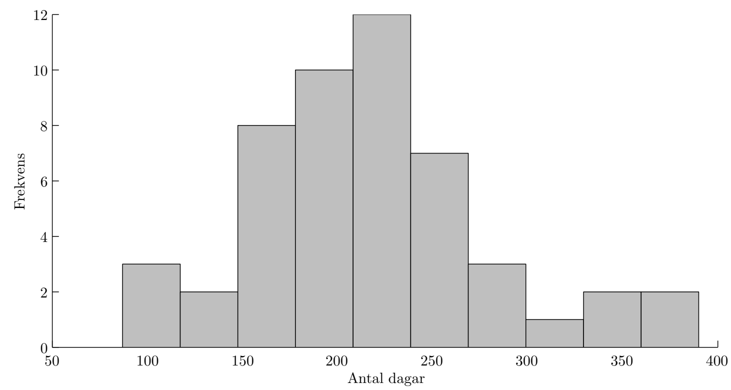
och stickprovsstandardavvikelse

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 64.25.$$



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.

11.7 De $n = 200$ observationerna $x_1, \ldots, x_n$ 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:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \approx \frac{1}{n}(17.5 \cdot 5 + 22.5 \cdot 19 + \cdots + 72.5 \cdot 4) = 34.225 \text{ tkr}$$

och stickprovsvariansen enligt:

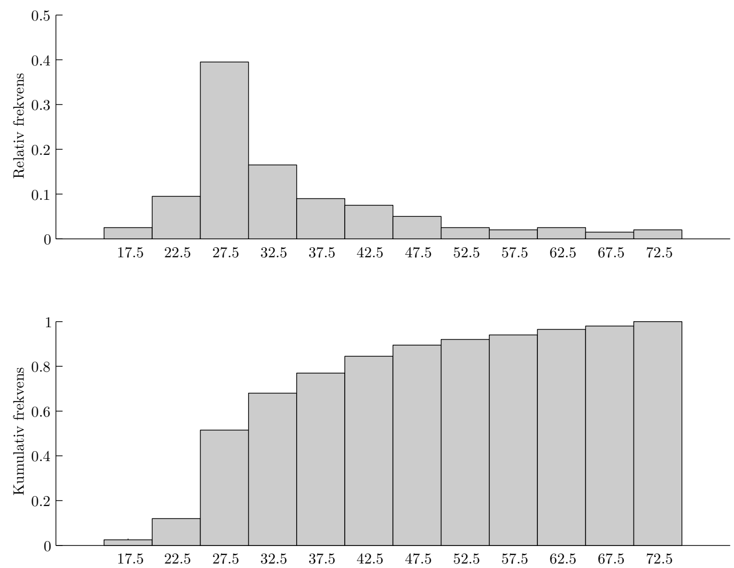
$$s_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) \approx \frac{1}{n-1}\left((17.5^2 \cdot 5 + 22.5^2 \cdot 19 + \cdots 72.5^2 \cdot 4) - n\overline{x}^2\right) = 139.09$$

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. Medelvärdet $\approx$ 34000 och medianen $\approx$ 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.

12.2 Låt $X_1, \ldots, X_n$ beskriva mätresultaten av batteriernas livslängder. Modell: $X_1, \ldots, X_n$ är oberoende och likafördelade med $m = E(X_i)$ och $\sigma^2 = V(X_i)$. Låt $x_1, \ldots, x_n$ vara utfallen på $X_1, \ldots, X_n$. Vi skattar $m = E(X_i)$ med

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 5.2$$

och σ^2 med

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right) = 1.7$$

så skattningen av σ är $s = \sqrt{1.7} = 1.3038$.

Vi beskriver skattningen $\bar{x}$ med stickprovsvariabeln

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

som har väntevärde

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = m$$

och varians

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n},$$

dvs $D\left(\overline{X}\right) = \sigma/\sqrt{n}$. Standardavvikelsen $D\left(\overline{X}\right) = \sigma/\sqrt{n}$ skattas med $s/\sqrt{n} = 0.5831$ (medelfelet för $\bar{x}$).

12.3 Om $X_1, \ldots, X_n$ är oberoende och N(m, σ), där $\sigma = 0.5$, så är $\overline{X}$ N$(m, \sigma/\sqrt{n})$. Alltså är

$$\begin{aligned} 0.99 &= P\left(|\overline{X} - m| \le 0.25\right) = P\left(-0.25 \le \overline{X} - m \le 0.25\right) = P\left(\frac{-0.25}{\sigma/\sqrt{n}} \le \frac{\overline{X} - m}{\sigma/\sqrt{n}} \le \frac{0.25}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{0.25}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-0.25}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{0.25}{\sigma/\sqrt{n}}\right) - \left(1 - \Phi\left(\frac{0.25}{\sigma/\sqrt{n}}\right)\right) = 2\Phi\left(\frac{0.25}{\sigma/\sqrt{n}}\right) - 1. \end{aligned}$$

Med $\sigma = 0.5$ får vi att

$$\Phi\left(\frac{0.25}{\sigma/\sqrt{n}}\right) = \frac{0.99+1}{2} = 0.995.$$

Ur tabeller för N$(0,1)$-fördelningsfunktionen $\Phi(\cdot)$ får man att

$$\frac{0.25}{\sigma/\sqrt{n}} = \lambda_{0.005} = 2.5758.$$

Alltså

$$\sqrt{n} = \frac{\lambda_{0.005}\,\sigma}{0.25} = \frac{2.5758 \cdot 0.5}{0.25} = 5.1516.$$

vilket ger $n = 26.5$ eller n minst lika med 27 (n heltal).

12.4 Låt $X_1, \ldots, X_n$ beskriva de oberoende mätresultaten med $E(X_i) = m$ och $V(X_i) = \sigma^2$. Observationer $x_1, \ldots, x_n$ ger skattningen $\overline{x}$ av m. Denna är väntevärdesriktig eftersom motsvarande stickprovsvariabel $\overline{X}$ har

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} m = m.$$

Tjebychevs olikhet säger att

$$P\left(|Y - E(Y)| > kD(Y)\right) \le \frac{1}{k^2},$$

vilket med $\epsilon = kD(Y)$ ger

$$P\left(|Y - E(Y)| > \epsilon\right) \le \frac{V(Y)}{\epsilon^2}.$$

Speciellt fås med $Y = \overline{X}$ att

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

så $D(Y) = \sigma/\sqrt{n}$. Alltså

$$P\left(|\overline{X} - m| > \epsilon\right) \le \frac{\sigma^2}{n\epsilon^2} \to 0$$

då $n \to \infty$ för varje val av $\epsilon > 0$ och skattningen $\overline{x}$ är konsistent.

12.5 Skattningar

$$\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_0}\right)$$

beskrivs av stickprovsvariabeln

$$\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{X_i}{x_0}\right)$$

där $X_1, \ldots, X_n$ är oberoende och Paretofördelade med parameter α, dvs.

$$F_X(x) = P(X \le x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha}$$

för $x > x_0$ och en konstant $\alpha > 0$. En paretofördelad stokastisk variabel har täthetsfunktion

$$f_X(x) = \frac{d}{dx}F_X(x) = x_0^{\alpha}\alpha x^{-(\alpha+1)},$$

för $x \geq x_0$. Alltså är väntevärdet

$$\begin{aligned}
E\left(\ln\left(\frac{X}{x_0}\right)\right) &= \int \ln(x/x_0)\, f_X(x)\, dx = \int_{x_0}^{\infty} \ln(x/x_0)\, x_0^{\alpha} \alpha x^{-(\alpha+1)}\, dx = \{u = x/x_0\} \\
&= \int_1^{\infty} \ln(u)\, \alpha u^{-(\alpha+1)}\, du = \left[\ln(u)\, \frac{\alpha u^{-\alpha}}{-\alpha}\right]_1^{\infty} + \frac{1}{\alpha}\int_1^{\infty} u^{-\alpha}\, \frac{1}{u}\, du \\
&= \left[\frac{u^{-\alpha}}{-\alpha}\right]_1^{\infty} = \frac{1}{\alpha}.
\end{aligned}$$

Vår skattningar är väntevärdesriktiga skattningar av $1/\alpha$ eftersom

$$E\left(\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{X_i}{x_0}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(\ln\left(\frac{X_i}{x_0}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\alpha} = \frac{1}{\alpha}.$$

12.6 Den stokastiska variabeln X är binomialfördelad, X är $\mathrm{Bin}(n, p)$ och Y är hypergeometriskt fördelad, Y är $\mathrm{Hyp}(N, n, p)$.

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

$$E(\widehat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{1}{n}\, np = p$$

och

$$\begin{aligned}
V(\widehat{p}) &= V\left(\frac{1}{n}Y\right) = \frac{1}{n^2}V(Y) = \frac{1}{n^2}\, np(1-p)\frac{N-n}{N-1} = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1} = V(p^*)\underbrace{\frac{N-n}{N-1}}_{<1} \\
&< V(p^*).
\end{aligned}$$

b) Skattningen $\widehat{p}$ är effektivast (har minst varians)

c) Med observationer erhålls skattningarna $p^* = \widehat{p} = 23/100 = 0.23$ av andelarna.

Vi skattar varianserna för skattningsvariablerna (stickprovsvariablerna) genom att sätta in skattningarna av p. Dvs, vi skattar

$$V(p^*) = \frac{p(1-p)}{n} \quad \text{med} \quad \frac{p^*(1-p^*)}{n} = 0.001771$$

och

$$V(\widehat{p}) = \frac{p(1-p)}{n}\,\frac{N-n}{N-1} \quad \text{med} \quad \frac{p^*(1-p^*)}{n}\,\frac{N-n}{N-1} = 0.00160.$$

Skattningarna av standardavvikelserna, 0.0421 resp. 0.0399, kallas skattningarnas medelfel.

12.7 Låt $X_1, \ldots, X_n$ beskriva mätfelen. Modell: $X_1, \ldots, X_n$ är oberoende och $\mathrm{N}(0, \sigma)$ med observerade värden $x_1, \ldots, x_n$. En skattning av σ ges av

$$\sigma_2^* = a_n \sum_{i=1}^{n} |x_i|$$

som beskrivs av

$$S_2 = a_n \sum_{i=1}^{n} |X_i|.$$

För $\mathrm{N}(0,\sigma)$-fördelad X är

$$\begin{aligned} E\left(|X|\right) &= \int x f_{|X|}(x)\,dx = \int_0^\infty x\,\frac{\sqrt{2}}{\sqrt{\pi}\sigma}\mathrm{e}^{-x^2/2\sigma^2}\,dx = \{\text{Subst. } u = x^2/2\} \\ &= \int_0^\infty \frac{\sqrt{2}}{\sqrt{\pi}\sigma}\mathrm{e}^{-u/\sigma^2}\,dx = \left[\frac{\sqrt{2}}{\sqrt{\pi}\sigma}\mathrm{e}^{-u/\sigma^2}(-\sigma^2)\right]_0^\infty = \sqrt{\frac{2}{\pi}}\sigma. \end{aligned}$$

Alltså är

$$E\left(S_2\right) = E\left(a_n \sum_{i=1}^{n}|X_i|\right) = a_n \sum_{i=1}^{n} E\left(|X_i|\right) = a_n \sum_{i=1}^{n}\sqrt{\frac{2}{\pi}}\sigma = a_n\sqrt{\frac{2}{\pi}}\sigma n.$$

Väljs

$$a_n = \frac{1}{n}\sqrt{\frac{\pi}{2}}$$

så är $E\left(S_2\right) = \sigma$ och skattningen σ_2^* är väntevärdesriktig. Utnyttjar vi det allmänna resultatet att

$$E\left(Y^2\right) = V\left(Y\right) + \left(E\left(Y\right)\right)^2$$

så får vi att

$$\begin{aligned} V\left(|X_i|\right) &= E\left(|X_i|^2\right) - \left(E\left(|X_i|\right)\right)^2 = E\left(X_i^2\right) - \left(E\left(|X_i|\right)\right)^2 = V\left(X_i\right) + \left(E\left(X_i\right)\right)^2 - \left(E\left(|X_i|\right)\right)^2 \\ &= \sigma^2 + 0^2 - \frac{2}{\pi}\sigma^2 = \frac{\pi-2}{\pi}\sigma^2 \end{aligned}$$

vilket ger

$$V\left(S_2\right) = V\left(a_n\sum_{i=1}^{n}|X_i|\right) = \{\text{oberoende}\} = a_n^2\sum_{i=1}^{n} V\left(|X_i|\right) = \left(\frac{1}{n}\sqrt{\frac{\pi}{2}}\right)^2 \sum_{i=1}^{n}\frac{\pi-2}{\pi}\sigma^2 = \frac{\pi-2}{2n}\sigma^2.$$

Den andra skattningen σ_1^*,

$$\sigma_1^* = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

beskrivs av

$$S_1 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$$

som är sådan att

$$E\left(S_1^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i^2\right) = \frac{1}{n}\sum_{i=1}^{n}\left(V\left(X_i\right) + \left(E\left(X_i\right)\right)^2\right) = \sigma^2.$$

Vidare så är

$$\begin{aligned} E\left(X^4\right) &= \int x^4\, f_X(x)\,dx = \int_{-\infty}^{\infty} x^4\,\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/2\sigma^2}\,dx = \{u = x^2/2\sigma\} \\ &= 2\cdot(2\sigma^2)^{3/2}\sigma\frac{1}{\sqrt{2\pi}}\underbrace{\int_0^\infty u^{3/2}\,\mathrm{e}^{-u}\,du}_{=\Gamma(5/2)} = \frac{4}{\sqrt{\pi}}\sigma^4\,\underbrace{\Gamma\left(\frac{5}{2}\right)}_{=\frac{3}{4}\sqrt{\pi}} = 3\sigma^4 \end{aligned}$$

vilket ger

$$\begin{aligned} V\left(S_1^2\right) &= V\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^n V\left(X_i^2\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n (E\left((X_i^2)^2\right) - (E\left(X_i^2\right))^2) = \frac{1}{n}(3\sigma^4 - (\sigma^2)^2) = \frac{2\sigma^4}{n}. \end{aligned}$$

Med funktionen $g(x) = \sqrt{x}$ så är $g'(x) = \frac{d}{dx}g(x) = \frac{1}{2\sqrt{x}}$ och med Gauss approximationsformler

$$\begin{aligned} E\left(S_1\right) = E\left(g(S_1^2)\right) &\approx g(E\left(S_1^2\right)) = g(\sigma^2) = \sqrt{\sigma^2} = \sigma \\ V\left(S_1\right) = V\left(g(S_1^2)\right) &\approx \left[g'(E\left(S_1^2\right))\right]^2 V\left(S_1^2\right) = [g'(\sigma^2)]^2\,\frac{2\sigma^4}{n} = \frac{1}{4\sigma^2}\,\frac{2\sigma^4}{n} = \frac{1}{2n}\sigma^2. \end{aligned}$$

Detta ger att

$$V\left(S_1\right) \approx \frac{1}{2n}\sigma^2 \leq \frac{\pi - 2}{2n}\sigma^2 = V\left(S_2\right)$$

och skattningen σ_1^* är effektivare än skattningen σ_2^*.

Med observationer $x_1, \ldots, x_5$:

$$0.17 \quad -0.26 \quad -0.42 \quad 0.18 \quad 0.02$$

fås skattningarna

$$\sigma_1^* = \sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2} = 0.24727 \qquad \sigma_2^* = \frac{1}{n}\sqrt{\frac{\pi}{2}}\sum_{i=1}^n |x_i| = 0.2632.$$

12.8 Låt $X_1, \ldots, X_n$ vara oberoende och $\mathrm{R}(0, \theta)$ med observerade värden $x_1, \ldots, x_n$. Skattningen

$$\theta_1^* = a_n\overline{x}$$

beskrivs av

$$\Theta_1^* = a_n\overline{X}$$

som har väntevärde

$$E\left(\Theta_1^*\right) = E\left(a_n\overline{X}\right) = a_n\frac{1}{n}\sum_{i=1}^n E\left(X_i\right) = a_n\,\frac{\theta}{2}.$$

Med $a_n = 2$ så är $E\left(\Theta_1^*\right) = \theta$ och skattningen θ_1^* är väntevärdesriktig. Med detta val så är

$$V\left(\Theta_1^*\right) = V\left(2\overline{X}\right) = 4V\left(\overline{X}\right) = 4\,\frac{V\left(X\right)}{n} = 4\,\frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.$$

Skattningen

$$\theta_2^* = b_n\ \max(x_1, \ldots, x_n)$$

beskrivs av

$$\Theta_2^* = b_n\ \max(X_1, \ldots, X_n).$$

Eftersom X_i är $\mathrm{R}(0, \theta)$ är $f_X(x) = 1/\theta$ för $0 \leq x \leq \theta$ och

$$P\left(X \leq x\right) = \int_0^x \frac{1}{\theta}\,dx = \frac{x}{\theta}.$$

för $0 \leq x \leq \theta$. Alltså är

$$P\left(\max(X_1, \ldots, X_n) \leq x\right) = P\left(X_1 \leq x, \ldots, X_n \leq x\right) = P\left(X_1 \leq x\right)\cdots P\left(X_n \leq x\right) = \left(\frac{x}{\theta}\right)^n.$$

Således är fördelningsfunktionen för Θ_2^*

$$F_{\Theta_2^*}(x) \quad = \quad P\left(\Theta_2^* \leq x\right) = P\left(b_n \max(X_1, \ldots, X_n) \leq x\right) = \frac{x^n}{b_n^n \theta^n}$$

och

$$f_{\Theta_2^*}(x) = \frac{d}{dx} F_{\Theta_2^*}(x) = n\,\frac{x^{n-1}}{b_n^n \theta^n}$$

för $0 \leq x \leq b_n \theta$. Alltså är Θ_2^* väntevärde

$$E\left(\Theta_2^*\right) \quad = \quad \int x f_{\Theta_2^*}(x)\,dx = \int_0^{b_n\theta} xn\,\frac{x^{n-1}}{b_n^n \theta^n}\,dx = \left[\frac{n}{b_n^n \theta^n (n+1)} x^{n+1}\right]_0^{b_n\theta} = \frac{n}{n+1} b_n \theta.$$

Med $b_n = (n+1)/n$ så är $E\left(\Theta_2^*\right) = \theta$ och skattningen θ_2^* är väntevärdesriktig. Vidare,

$$E\left((\Theta_2^*)^2\right) \quad = \quad \int x^2 f_{\Theta_2^*}(x)\,dx = \int_0^{b_n\theta} x^2 n\,\frac{x^{n-1}}{b_n^n \theta^n}\,dx = \left[\frac{n}{b_n^n \theta^n (n+2)} x^{n+2}\right]_0^{b_n\theta} = \frac{n}{n+2} b_n^2 \theta^2$$

så

$$V\left(\Theta_2^*\right) = E\left((\Theta_2^*)^2\right) - \left(E\left(\Theta_2^*\right)\right)^2 = \frac{n}{n+2}\,\frac{(n+1)^2}{n^2}\,\theta^2 - \theta^2 = \frac{\theta^2}{n(n+2)}.$$

Alltså är

$$V\left(\Theta_2^*\right) = \frac{\theta^2}{n(n+2)} \leq \frac{\theta^2}{3n} = V\left(\Theta_1^*\right)$$

(med likhet endast om $n = 1$) och skattningen θ_2^* är effektivare än θ_1^*.

Med observationer $x_1, \ldots, x_5$:

$$0.3 \quad 0.8 \quad 1.9 \quad 0.2 \quad 1.5$$

fås skattningarna

$$\theta_1^* = 2\overline{x} = 1.88 \qquad \theta_2^* = \frac{6}{5}\max(x_1, \ldots, x_n) = 2.28.$$

Notera att $\theta_1^* < x_3$.

12.9 I en sjö finns N fiskar. Man fångar och märker a stycken av dessa. Vid en tid senare fångar man n fiskar och låter X beskriva antalet fiskar av dessa som är märkta. Modell: X är hypergeometriskt fördelat, dvs

$$P\left(X = k\right) = \frac{\binom{a}{k}\binom{N-a}{n-k}}{\binom{N}{n}}$$

som har

$$E\left(X\right) = n\frac{a}{N} \quad \text{och} \quad V\left(X\right) = n\frac{a}{N}\,\frac{N-a}{N}\,\frac{N-n}{N-1}.$$

Med observerat värde x är en skattning av N

$$n^* = \frac{na}{x}$$

som beskrivs av stickprovsvariabeln

$$N^* = \frac{na}{X} = g(X)$$

där funktionen $g(x) = na/x$ har derivata

$$g'(x) = \frac{d}{dx} g(x) = -\frac{na}{x^2}.$$

Med Gauss approximationsformler har vi att

$$\begin{aligned}
E(N^*) = E(g(X)) &\approx g(E(X)) = \frac{na}{na/N} = N \\
V(N^*) = V(g(X)) &\approx [g'(E(X))]^2 V(X) = [g'(na/N)]^2\, n\frac{a}{N}\,\frac{N-a}{N}\,\frac{N-n}{N-1} \\
&= \frac{N^2(N-a)(N-n)}{na(N-1)}.
\end{aligned}$$

12.10 Låt λ vara intensiteten av partiklar under en tidsperiod av längd t och X beskriva antalet registrerade partiklar. Modell: X är Poissonfördelad med parameter λt. Då är

$$E(X) = \lambda t \qquad V(X) = \lambda t$$

och $\lambda = E(X)/t$. Med utfallet x på X ges en skattning λ^* av λ av

$$\lambda^* = x/t$$

som beskrivs av $\Lambda^* = X/t$. Då är

$$E(\Lambda^*) = E\left(\frac{X}{t}\right) = \frac{1}{t}E(X) = \lambda$$

och

$$V(\Lambda^*) = V\left(\frac{X}{t}\right) = \frac{1}{t^2}V(X) = \frac{\lambda}{t}.$$

Tjebychevs olikhet säger att

$$P(|Y - E(Y)| > kD(Y)) \leq \frac{1}{k^2},$$

vilket med $\epsilon = kD(Y)$ ger

$$P(|Y - E(Y)| > \epsilon) \leq \frac{V(Y)}{\epsilon^2}.$$

Speciellt fås med $Y = \Lambda^*$ att

$$P(|\Lambda^* - \lambda| > \epsilon) \leq \frac{\lambda}{t\epsilon^2} \to 0$$

då $t \to \infty$ för varje val av $\epsilon > 0$ och skattningen λ^* är konsistent.

12.11 Att θ_1^* och θ_2^* är väntevärdesriktiga skattningsvariabler av en parameter θ betyder att $E(\theta_1^*) = \theta$ och $E(\theta_2^*) = \theta$. Med

$$\theta^* = a\theta_1^* + (1-a)\theta_2^*$$

så är

$$E(\theta^*) = E(a\theta_1^* + (1-a)\theta_2^*) = a\underbrace{E(\theta_1^*)}_{=\theta} + (1-a)\underbrace{E(\theta_2^*)}_{=\theta} = a\theta + (1-a)\theta = \theta$$

för alla värden a. Alltså är θ^* väntevärdesriktig. Vidare så är

$$\begin{aligned}
V(\theta^*) &= V(a\theta_1^* + (1-a)\theta_2^*) = \{\text{oberoende}\} = a^2\underbrace{V(\theta_1^*)}_{=\sigma_1^2} + (1-a)^2\underbrace{V(\theta_2^*)}_{=\sigma_2^2} \\
&= a^2\sigma_1^2 + (1-a)^2\sigma_2^2 = g(a),
\end{aligned}$$

dvs någon funktion av a. Vi söker den effektivaste skattningen, dvs det värde på a som gör variansen så liten som möjligt. Vi minimerar $g(a)$ med avseende på a genom att lösa

$$\frac{d}{da}g(a) = \sigma_1^2 \cdot 2a + \sigma_2^2 \cdot 2(1-a)(-1) = 2\left[a(\sigma_1^2 + \sigma_2^2) - \sigma_2^2\right] = 0$$

med lösningen

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Teckenstudium av andraderivatan visar att det är ett minimum som vi fått fram.

12.12 Låt $X_1, \ldots, X_n$ beskriva antalet samtal under de olika dagar. Modell: $X_1, \ldots, X_n$ är oberoende Poissonfördelad med parameter m. Låt $x_1, \ldots, x_n$ vara utfall av $X_1, \ldots, X_n$.

Maximum likelihoodskattningen av m är det värde på m som maximerar

$$\begin{aligned} L(m) &= P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= \frac{m^{x_1}}{x_1!}\mathrm{e}^{-m} \cdots \frac{m^{x_n}}{x_n!}\mathrm{e}^{-m} = \frac{m^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!}\mathrm{e}^{-nm}. \end{aligned}$$

Det är samma m som maximerar

$$\ln(L(m)) = \ln\left(\frac{m^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!}\mathrm{e}^{-nm}\right) = \ln(m)\sum_{i=1}^n x_i - nm - \sum_{i=1}^n \ln(x_i!).$$

Lösning av

$$0 = \frac{d}{dm}\ln(L(m)) = \frac{d}{dm}\left[\ln(m)\sum_{i=1}^n x_i - nm - \sum_{i=1}^n \ln(x_i!)\right] = \frac{1}{m}\sum_{i=1}^n x_i - n$$

ger $m = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$. Kontroll av andraderivatan ger att detta är ett maximum. Skattningen $m^* = \overline{x}$ av m beskrivs av $\overline{X}$ som har

$$E\left(\overline{X}\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\sum_{i=1}^n m = m$$

och

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^n X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^n \underbrace{V(X_i)}_{=m} = \frac{m}{n},$$

dvs $D\left(\overline{X}\right) = \sqrt{m/n}$.

Med observationer $x_1, \ldots, x_8$:

$$115 \quad 82 \quad 108 \quad 106 \quad 118 \quad 87 \quad 99 \quad 92$$

fås skattningen $m^* = \overline{x} = 100.88$. En skattning av $D\left(\overline{X}\right) = \sqrt{m/n}$ ges av

$$\sqrt{\frac{m^*}{n}} = \sqrt{\frac{100.88}{8}} = 3.551$$

och kallas skattningens medelfel.

12.13 Låt $X_1, \ldots, X_n$ beskriva antalet registrerade partiklar under tidsintervall av längd $t_1, \ldots, t_n$ med observerade värden $x_1, \ldots, x_n$. Modell: $X_1, \ldots, X_n$ är oberoende och X_i är Poissonfördelad med parameter (väntevärde) λt_i.

Maximum likelihoodskattningen av λ är det värde på λ som maximerar

$$\begin{aligned} L(\lambda) &= P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= \frac{(\lambda t_1)^{x_1}}{x_1!}\mathrm{e}^{-\lambda t_1} \cdots \frac{(\lambda t_n)^{x_n}}{x_n!}\mathrm{e}^{-\lambda t_n} = \frac{\lambda^{\sum_{i=1}^n x_i} t_1^{x_1} \cdots t_n^{x_n}}{x_1! \cdots x_n!}\mathrm{e}^{-\lambda \sum_{i=1}^n t_i}. \end{aligned}$$

Det är samma λ som maximerar

$$\ln(L(\lambda)) \quad = \quad \ln\left(\frac{\lambda^{\sum_{i=1}^{n} x_i} t_1^{x_1}\cdots t_n^{x_n}}{x_1!\cdots x_n!} \mathrm{e}^{-\lambda \sum_{i=1}^{n} t_i}\right) = \ln(\lambda)\sum_{i=1}^{n} x_i - \lambda \sum_{i=1}^{n} t_i + \sum_{i=1}^{n} \ln\left(t_i^{x_i}/x_i!\right).$$

Lösning av

$$0 = \frac{d}{d\lambda}\ln(L(\lambda)) = \frac{d}{d\lambda}\left[\ln(\lambda)\sum_{i=1}^{n} x_i - \lambda\sum_{i=1}^{n} t_i + \sum_{i=1}^{n}\ln\left(t_i^{x_i}/x_i!\right)\right] = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} t_i$$

ger

$$\lambda = \sum_{i=1}^{n} x_i \Big/ \sum_{i=1}^{n} t_i.$$

Kontroll av andraderivatan ger att detta är ett maximum.

Med observationer $x_1, \ldots, x_3$ på intervall av längd $t_1 = 10$, $t_2 = 10$ och $t_3 = 50$:

$$122 \quad 137 \quad 761$$

fås skattningen $\lambda^* = (122 + 137 + 761)/(10 + 10 + 50) = 14.571$ partiklar/minut.

12.14 Exponentialfördelningen har täthetsfunktion $f_X(x) = \frac{1}{m}\mathrm{e}^{-x/m}$ för $x \geq 0$.

a) Likelihoodfunktionen blir därför

$$L(m) = f_X(x_1)\cdots f_X(x_n) = \frac{1}{m}\mathrm{e}^{-x_1/m}\cdots\frac{1}{m}\mathrm{e}^{-x_n/m} = \frac{1}{m^n}\mathrm{e}^{-\frac{1}{m}\sum_{i=1}^{n} x_i}.$$

Maximum-likelihoodskattningen av m är det värde på m som maximerar $L(m)$. Det är samma m som maximerar $\ln(L(m))$ så vi maximerar

$$\ln(L(m)) = -n\ln(m) - \frac{1}{m}\sum_{i=1}^{n} x_i.$$

Funktionen deriveras och

$$\frac{d}{dm}\ln(L(m)) = -n\frac{1}{m} + \frac{1}{m^2}\sum_{i=1}^{n} x_i$$

som är 0 för $m = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$. Alltså, m skattas med $m^* = \overline{x}$.

b) Skattningen $\overline{x}$ beskrivs av skattningsvariabeln (stickprovsvariabeln) $m^* = \overline{X}$. Vi erhåller

$$E\left(m^*\right) = E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{E\left(X_i\right)}_{=m} = m$$

dvs skattningen är väntevärdesriktig, och

$$V\left(m^*\right) = V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{V\left(X_i\right)}_{=m^2} = \frac{m^2}{n}.$$

c) Vi räknar som att alla observationer i ett intervall hamnat i mittpunkten. Det ger oss att

$$m^* = \overline{x} \approx \frac{27\cdot 12.5 + 18\cdot 37.5 + \cdots + 2\cdot 13.5}{80} = \frac{3775}{80} \approx 47.2.$$

12.15 Låt $X_1, \ldots, X_n$ beskriva tiderna mellan fel i den komplicerade tekniska utrustningen, med observerade värden $x_1, \ldots, x_n$. Modell: $X_1, \ldots, X_n$ är oberoende och exponentialfördelade med parameter m.

Maximum likelihoodskattningen av m är det värde på m som maximerar

$$\begin{aligned} L(m) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \frac{1}{m}\mathrm{e}^{-x_1/m} \cdots \frac{1}{m}\mathrm{e}^{-x_n/m} = \frac{1}{m^n}\mathrm{e}^{-\frac{1}{m}\sum_{i=1}^n x_i}. \end{aligned}$$

Det är samma m som maximerar

$$\ln(L(m)) = \ln\left(\frac{1}{m^n}\mathrm{e}^{-\frac{1}{m}\sum_{i=1}^n x_i}\right) = -n\ln(m) - \frac{1}{m}\sum_{i=1}^n x_i.$$

Lösning av

$$0 = \frac{d}{dm}\ln(L(m)) = \frac{d}{dm}\left[-n\ln(m) - \frac{1}{m}\sum_{i=1}^n x_i\right] = -n\frac{1}{m} + \frac{1}{m^2}\sum_{i=1}^n x_i = \frac{1}{m}\left[-n + \frac{1}{m}\sum_{i=1}^n x_i\right]$$

ger $m = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$. Kontroll av andraderivatan ger att detta är ett maximum. Skattningen $m^* = \overline{x}$ av m beskrivs av $\overline{X}$ som har

$$E\left(\overline{X}\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\sum_{i=1}^n m = m$$

så skattningen m^* är väntevärdesriktig. Vidare

$$V\left(\overline{X}\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^n \underbrace{E(X_i)}_{=m^2} = \frac{m^2}{n},$$

dvs $D\left(\overline{X}\right) = m/\sqrt{n}$.

Med observationer $x_1, \ldots, x_n$ givna av frekvenstabellen:

Klass (tim)	Klassmitt	Frekvens
0−25	12.5	27
25−50	37.5	18
50−75	62.5	20
75−100	87.5	9
100−125	112.5	4
125−150	137.5	2

Medelvärdet beräknas enligt:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i \approx \frac{1}{n}(12.5 \cdot 27 + 37.5 \cdot 18 + \cdots + 137.5 \cdot 2) = 47.188 \text{ tim}$$

Anmärkning: Med klassindelade data skulle uppgiften lösas på följande sätt. Låt $a_0, a_1, \ldots, a_6$ vara klassgränserna $a_i = 25i$, $i = 0, \ldots, 6$, $a_7 = \infty$, och $y_0, \ldots, y_6$ som antalet observationer i på intervall (a_i, a_{i+1}), $i = 0, \ldots, 6$.

Då är sannolikheten att få en observation i kategori i

$$\begin{aligned} p_i &= P(a_i < X \le a_{i+1}) = F_X(a_{i+1}) - F_X(a_i) = 1 - \mathrm{e}^{-a_{i+1}/m} - (1 - \mathrm{e}^{-a_i/m}) \\ &= \mathrm{e}^{-25i/m} - \mathrm{e}^{-25(i+1)/m} = \mathrm{e}^{-25i/m}(1 - \mathrm{e}^{-25/m}) = p^i(1-p) \end{aligned}$$

där $p = \mathrm{e}^{-25/m}$.

Klass (tim)	Frekvens, y_i	$P(X \in \text{klass})$
$(0, 25]$	27	$p_0 = 1 - \mathrm{e}^{-25/m} = 1 - p$
$(25, 50]$	18	$p_1 = \mathrm{e}^{-25/m} - \mathrm{e}^{50/m} = p(1-p)$
$\vdots$	$\vdots$	$\vdots$
$(125, 150]$	2	$p_5 = \mathrm{e}^{-125/m} - \mathrm{e}^{150/m} = p^5(1-p)$
$(150, \infty)$	0	$p_6 = \mathrm{e}^{-150/m} = p^6$

Maximum likelihoodskattningen av p är det värde på p som maximerar

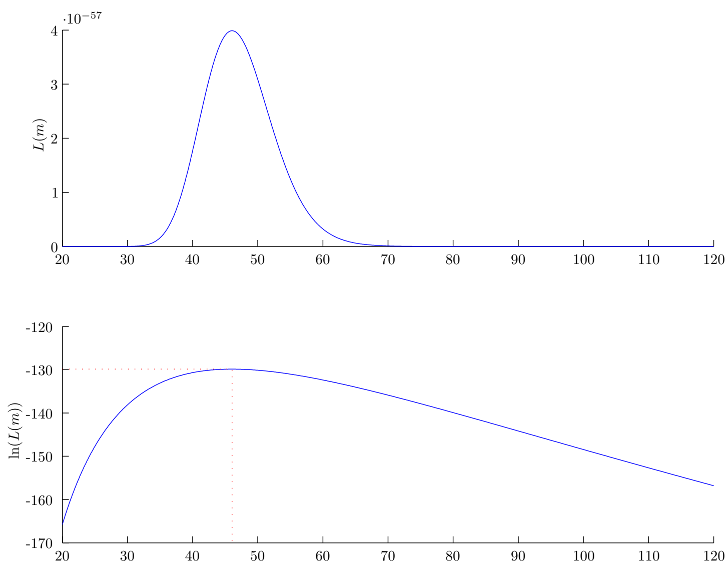
$$\begin{aligned} L(p) &= P(Y_0 = y_0, \dots, Y_6 = y_6) = \frac{(\sum y_i)!}{y_0! y_1! \cdots y_6!} p_0^{y_0} p_1^{y_1} \cdots p_6^{y_6} \\ &= C\,(1-p)^{y_1} \cdot p^{y_1}(1-p)^{y_2} \cdot p^{2y_2}(1-p)^{y_2} \cdots p^{6y_6}(1-p)^{y_6} \\ &= C\, p^{\sum_{k=0}^{6} k y_k} (1-p)^{\sum_{k=0}^{6} y_k} = C\, p^a (1-p)^b \end{aligned}$$

med $C = \frac{(\sum y_i)!}{y_0! y_1! \cdots y_6!}$, $a = \sum_{k=0}^{6} k y_k = 111$ och $b = \sum_{k=0}^{6} y_k = 80$. Lösning av

$$\frac{d}{dp} L(p) = C\, a p^{a-1}(1-p)^b - p^a b (1-p)^b = C\, p^{a-1}(1-p)^{b-1} \underbrace{[ap - b(1-p)]}_{=0} = 0$$

ger $p^* = b/(a+b) = 111/191$. Skattningen p^* ger en skattning m^* eftersom

$$\mathrm{e}^{-25/m^*} = p^* \quad \text{så fås} \quad m^* = -25/\ln(p^*) = -25/\ln(111/191) = 46.0623.$$



Likelihoodfunktionen $L(m)$ överst och $\ln(L(m))$ nederst. Maximum till båda erhålls i punkten $m = 46.062$.

12.16 Låt $X_1, \dots, X_n$ vara oberoende $\mathrm{R}[0, \theta]$ med observerade värden $x_1, \dots, x_n$. Tätheten för de stokastiska variablerna är

$$f_{X_i}(x) = \frac{1}{\theta} \qquad \text{för } 0 \le x \le \theta.$$

Maximum likelihoodskattningen är det värde på θ som maximerar

$$L(\theta) = f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \{\text{oberoende}\} = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{1}{\theta} = \theta^{-n}$$

för $0 \leq x_1, \ldots, x_n \leq \theta$. Eftersom $L(\theta)$ är avtagande i θ skall man göra θ så liten som möjligt. Kravet $0 \leq x_1, \ldots, x_n \leq \theta$ gör att det minsta möjliga värdet på θ ges av $\theta = \max(x_1, \ldots, x_n)$.

12.17 Låt $X_1, \ldots, X_n$ vara oberoende R$[-\theta, \theta]$ med observerade värden $x_1, \ldots, x_n$. Tätheten för de stokastiska variablerna är

$$f_{X_i}(x) = \frac{1}{2\theta} \qquad \text{för } -\theta \leq x \leq \theta.$$

Maximum likelihoodskattningen är det värde på θ som maximerar

$$L(\theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \{\text{oberoende}\} = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{1}{2\theta} = (2\theta)^{-n}$$

för $-\theta \leq x_1, \ldots, x_n \leq \theta$. Eftersom $L(\theta)$ är avtagande i θ skall man göra θ så liten som möjligt. Kravet $-\theta \leq x_1, \ldots, x_n \leq \theta$, dvs. $|x_1|, \ldots, |x_n| \leq \theta$, gör att det minsta möjliga värdet på θ ges av $\theta = \max(|x_1|, \ldots, |x_n|)$.

12.18 Låt X beskriva livslängden för en komponent. Modell: X är exponentialfördelad med väntevärde m, dvs

$$f_X(x) = \frac{1}{m}\mathrm{e}^{-x/m} \quad x \geq 0.$$

Då är

$$p = P(X > t) = \int_t^{\infty} f_X(x)\,dx = \mathrm{e}^{-t/m}$$

sannolikheten att en komponent fungerar tiden t. Av n oberoende komponenter låt Y beskriva antalet som fungerar vid tiden t. Då är Y binomialfördelad, Y är Bin(n, p). Med utfallet k på Y så är Maximum likelihoodskattningen av m det värde på m som maximerar

$$L(m) = P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} (\mathrm{e}^{-t/m})^k (1 - (\mathrm{e}^{-t/m}))^{n-k}.$$

Det är samma värde på m som maximerar

$$\log(L(m)) = \log\left(\binom{n}{k}(\mathrm{e}^{-t/m})^k(1-(\mathrm{e}^{-t/m}))^{n-k}\right) = \log\binom{n}{k} - \frac{tk}{m} + (n-k)\log\left(1 - \mathrm{e}^{-t/m}\right).$$

Lösning av

$$\begin{aligned} 0 &= \frac{d}{dm}\log(L(m)) = \frac{d}{dm}\left[\log\binom{n}{k} - \frac{tk}{m} + (n-k)\log\left(1-\mathrm{e}^{-t/m}\right)\right] = \frac{tk}{m^2} + \frac{-t(n-k)}{m^2(\mathrm{e}^{t/m}-1)} \\ &= \frac{t}{m^2}\underbrace{\left[k + \frac{-(n-k)}{\mathrm{e}^{t/m}-1}\right]}_{=0} \end{aligned}$$

ger

$$m = \frac{t}{\log(n/k)}.$$

Att av $n = 100$ komponenter har $k = 90$ gått sönder vid tiden $t = 1000$ ger oss skattningen

$$m^* = \frac{t}{\log(n/k)} = \frac{1000}{\log(100/90)} = 9491.2 \text{ timmar.}$$

Notera att skattningen p^* av p

$$p^* = k/n = 90/100 = 0.90$$

säger att 90% av komponenterna håller denna tid. Vi vet att $p = \mathrm{e}^{-t/m}$ så

$$p^* = \mathrm{e}^{-t/m^*} \quad \text{ger} \quad m^* = -t/\log(p^*) = 9491.2.$$

12.19 Låt X beskriva antalet händelser under ett tidsintervall av enhetslängd. Modell: X är Poissonfördelad med parameter m. Då är

$$p = P(X = 0) = \frac{m^0}{0!}\mathrm{e}^{-m} = \mathrm{e}^{-m}$$

sannolikheten att under ett tidsintervall inte finns någon händelse. Av n oberoende tidsintervall låt Y beskriva antalet som inte rymmer någon händelse. Då är Y binomialfördelad, Y är $\mathrm{Bin}(n,p)$. Med utfallet k på Y så är Maximum likelihoodskattningen av m det värde på m som maximerar

$$L(m) = P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} (\mathrm{e}^{-m})^k (1 - (\mathrm{e}^{-m}))^{n-k}.$$

Det är samma värde på m som maximerar

$$\log(L(m)) = \log\left(\binom{n}{k}(\mathrm{e}^{-m})^k(1-(\mathrm{e}^{-m}))^{n-k}\right) = \log\binom{n}{k} - mk + (n-k)\log\left(1-\mathrm{e}^{-m}\right).$$

Lösning av

$$0 = \frac{d}{dm}\log(L(m)) = \frac{d}{dm}\left[\log\binom{n}{k} - mk + (n-k)\log\left(1-\mathrm{e}^{-m}\right)\right] = -k + \frac{n-k}{\mathrm{e}^m - 1}$$

ger

$$m = \ln(n/k).$$

Att av $n = 100$ tidsintervall fanns $k = 82$ stycken som inte rymde någon händelse ger oss skattningen

$$m^* = \log(n/k) = 0.19845.$$

Notera att skattningen p^* av p

$$p^* = k/n = 82/100 = 0.82$$

säger att 82% av de observerade tidsintervallen inte rymmer någon händelse. Vi vet att $p = \mathrm{e}^{-m}$ så

$$p^* = \mathrm{e}^{-m^*} \quad \text{ger} \quad m^* = -\log(p^*) = 0.19845.$$

12.20 Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(m,\sigma)$ med observerade värden $x_1, \ldots, x_n$. Tätheten för de stokastiska variablerna är

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-m)^2/2\sigma^2}.$$

Maximum likelihoodskattningen av (m,σ) är det värde på (m,σ) som maximerar

$$\begin{aligned} L(m,\sigma) &= f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \{\text{oberoende}\} = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x_i-m)^2/2\sigma^2} \\ &= (2\pi)^{-n/2}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-m)^2\right\}. \end{aligned}$$

Detta värde är samma värde som maximerar $\ln L(m,\sigma)$ där

$$\ln(L(m,\sigma)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-m)^2.$$

Derivering med avseende på m ger

$$\frac{\partial}{\partial m}\ln L(m,\sigma) = -0 - 0 - \frac{1}{2\sigma^2}\sum_{i=1}^n 2(x_i-m)(-1) = \frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i - nm\right)$$

som, satt till noll, ger lösningen

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

för alla värden på σ. Insatt i $L(m, \sigma)$ och derivering med avseende på σ ger

$$\frac{\partial}{\partial \sigma} \ln L(\overline{x}, \sigma) \quad = \quad -0 - n\frac{1}{\sigma} + \frac{2}{\sigma^3} \cdot \frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{\sigma}\left(-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)$$

dvs, satt till noll, ger lösningen

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Alltså, skattningarna av m och σ^2 ges av $m^* = \overline{x}$ och $(\sigma^2)^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Denna skattning är inte väntevärdesriktig eftersom motsvarande stickprovsvariabel (skattningsvariabel) har väntevärde

$$\begin{aligned} E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) &= E\left(\frac{n-1}{n}\,\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\underbrace{E\left(S^2\right)}_{=\sigma^2} \\ &= \frac{n-1}{n}\sigma^2 \neq \sigma^2. \end{aligned}$$

Genom att multiplicera med konstanten $n/(n-1)$ så är $n/(n-1)(\sigma^2)^* = s^2$ och alltså väntevärdesriktig.

12.21 Låt $X_1, \ldots, X_n$ vara oberoende $\Gamma(a, p)$-fördelad stokastiska variabler med observerade värden $x_1, \ldots, x_n$.

Maximum likelihoodskattningen av (a, p) är det värde på (a, p) som maximerar

$$\begin{aligned} L(a, p) &= f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \frac{1}{a^p \Gamma(p)} x_1^{p-1} \mathrm{e}^{-x_1/a} \cdots \frac{1}{a^p \Gamma(p)} x_n^{p-1} \mathrm{e}^{-x_n/a} = \frac{(x_1 \cdots x_n)^{p-1}}{a^{np}\Gamma^n(p)} \mathrm{e}^{-\frac{1}{a}\sum_{i=1}^{n} x_i}. \end{aligned}$$

Det är samma (a, p) som maximerar

$$\begin{aligned} \ln(L(a, p)) &= \ln\left(\frac{(x_1 \cdots x_n)^{p-1}}{a^{np}\Gamma^n(p)} \mathrm{e}^{-\frac{1}{a}\sum_{i=1}^{n} x_i}\right) \\ &= (p-1)\sum_{i=1}^{n}\ln(x_i) - np\ln(a) - n\ln(\Gamma(p)) - \frac{1}{a}\sum_{i=1}^{n} x_i. \end{aligned}$$

Derivering med avseende på a och p ger

$$\begin{aligned} \frac{\partial}{\partial a}\ln(L(a, p)) &= -np\frac{1}{a} + \frac{1}{a^2}\sum_{i=1}^{n} x_i = \frac{1}{a}\left[-np + \frac{1}{a}\sum_{i=1}^{n} x_i\right] \\ \frac{\partial}{\partial p}\ln(L(a, p)) &= \sum_{i=1}^{n}\ln(x_i) - n\ln(a) - n\frac{1}{\Gamma(p)}\Gamma'(p) \end{aligned}$$

Sätts derivatorna till 0 fås ekvationssystemet att skattningarna a^* och p^* uppfyller

$$\left\{ \begin{aligned} a^* &= \frac{1}{np^*}\sum_{i=1}^{n} x_i = \frac{\overline{x}}{p^*} \\ \frac{\Gamma'(p^*)}{\Gamma(p^*)} &= \frac{1}{n}\sum_{i=1}^{n}\ln(x_i) - \ln(a^*) \end{aligned} \right.$$

12.22 Låt X vara en fördröjd exponentialfördelning, dvs.

$$f_X(x) = \frac{1}{m}\mathrm{e}^{-(x-a)/m}, \quad x \geq a.$$

(Detta är fördelningen för $X'+a$ där X' är exponentialfördelad(m).) Uppgiften är att baserat på oberoende observationer $x_1, \ldots, x_n$ av X skatta m och a.

Maximum likelihoodskattningen av (a, m) är det värde på (a, m) som maximerar

$$\begin{aligned} L(a,m) &= f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) \\ &= \frac{1}{m}\mathrm{e}^{-(x_1-a)/m}\cdots\frac{1}{m}\mathrm{e}^{-(x_n-a)/m} = \frac{1}{m^n}\mathrm{e}^{-\frac{1}{m}\sum_{i=1}^n x_i}\mathrm{e}^{na/m} \end{aligned}$$

Funktionen $L(a, m)$ är växande i a. Kravet $a \leq x_1, \ldots, x_n$ gör att det största a som kan väljas är $a = \min(x_1, \ldots, x_n)$.

Logaritmering ger

$$\ln(L(a,m)) = \ln\left(\frac{1}{m^n}\mathrm{e}^{-\frac{1}{m}\sum_{i=1}^n x_i}\mathrm{e}^{na/m}\right) = -n\ln(m) - \frac{1}{m}\sum_{i=1}^n x_i + \frac{na}{m}.$$

Derivering med avseende på m ger

$$\frac{\partial}{\partial m}\ln(L(a,p)) = -n\frac{1}{m} + \frac{1}{m^2}\sum_{i=1}^n x_i - \frac{na}{m^2} = \frac{1}{m}\left[-n + \frac{1}{m}\sum_{i=1}^n (x_i - a)\right].$$

Sätts derivatan till 0 fås

$$m = \frac{1}{n}\sum_{i=1}^n (x_i - a) = \overline{x} - a.$$

Med observationer $x_1, \ldots, x_{10}$:

7	9	12	8	14	9	17	8	13	10

fås skattningarna

$$a^* = \min(x_1,\ldots,x_n) = 7 \qquad m^* = \overline{x} - a^* = 10.7 - 7 = 3.7.$$

12.23 En kvadrat med arean θ har sidlängd $\sqrt{\theta}$. Låt X beskriva en sidlängdsmätning utan systematiska fel. Modell: $E(X) = \sqrt{\theta}$ och $V(X) = \sigma^2$. Låt $X_1, \ldots, X_n$ beskriva oberoende sidlängsmätningar med observerade värden $x_1, \ldots, x_n$.

Minsta kvadratmetodens skattning av θ är det värde på θ som minimerar

$$Q(\theta) = \sum_{i=1}^n (x_i - E(X_i))^2 = \sum_{i=1}^n (x_i - \sqrt{\theta})^2.$$

Lösning av

$$0 = \frac{d}{d\theta}Q(\theta) = \frac{d}{d\theta}\left[\sum_{i=1}^n (x_i - \sqrt{\theta})^2\right] = \sum_{i=1}^n 2(x_i - \sqrt{\theta})\frac{-1}{2\sqrt{\theta}} = \frac{-1}{\sqrt{\theta}}\underbrace{\left[\sum_{i=1}^n x_i - n\sqrt{\theta}\right]}_{=0}$$

ger

$$\theta = \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2 = \overline{x}^2.$$

Kontroll av andraderivatan ger att detta är ett minimum. Skattningen

$$\theta^* = \overline{x}^2$$

beskrivs av

$$\Theta^* = \overline{X}^2$$

som har

$$E(\Theta^*) = E\left(\overline{X}^2\right) = V\left(\overline{X}\right) + \left(E\left(\overline{X}\right)\right)^2 = \frac{\sigma^2}{n} + (\sqrt{\theta})^2 = \underbrace{\frac{\sigma^2}{n}}_{>0} + \theta > \theta$$

så skattningen θ^* är inte väntevärdesriktig.

12.24 Låt $X_1, \dots, X_n$ vara oberoende $\mathrm{N}(m_i(\theta), \omega k_i)$-fördelade stokastiska variabler med observerade värden $x_1, \dots, x_n$.

Minsta kvadratmetodens skattning av θ är det värde på θ som minimerar

$$Q(\theta) = \sum_{i=1}^{n} \frac{1}{V(X_i)} (x_i - E(X_i))^2 = \sum_{i=1}^{n} \frac{1}{\omega^2 k_i^2} (x_i - m_i(\theta))^2.$$

Maximum likelihoodmetodens skattning av θ är det värde som på θ som maximerar

$$\begin{aligned} L(\theta) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} k_i \omega} \mathrm{e}^{-(x_i - m_i(\theta))^2 / 2k_i^2 \omega^2} \\ &= \underbrace{\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} k_i \omega} \right)}_{=C(\omega)} \exp\left\{ -\frac{1}{2k_i^2 \omega^2} \sum_{i=1}^{n} (x_i - m_i(\theta))^2 \right\} = C(\omega) \exp^{-\frac{1}{2} Q(\theta)}. \end{aligned}$$

Att maximera

$$C(\omega) \exp^{-\frac{1}{2} Q(\theta)}$$

är samma sak som att minimera $Q(\theta)$. Alltså ger maximum likelihoodmetoden och minsta kvadratmetoden samma skattningar av θ.

12.25 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

$$E(X_1) = E(X_2) = E(X_3) = \theta_1 + \theta_2 \qquad E(X_4) = E(X_5) = \theta_1$$

och $V(X_i) = \sigma^2$. Minsta kvadratmetodens skattning av (θ_1, θ_2) är det värde på (θ_1, θ_2) som minimerar

$$\begin{aligned} Q(\theta_1, \theta_2) = \sum_{i=1}^{n} (x_i - E(X_i))^2 &= (x_1 - (\theta_1 + \theta_2))^2 + (x_2 - (\theta_1 + \theta_2))^2 + (x_3 - (\theta_1 + \theta_2))^2 \\ &\quad + (x_4 - \theta_1)^2 + (x_5 - \theta_1)^2. \end{aligned}$$

Derivering med avseende på θ_1 och θ_2 ger

$$\begin{aligned} \frac{\partial}{\partial \theta_1} Q(\theta_1, \theta_2) &= -2(x_1 - (\theta_1 + \theta_2)) - 2(x_2 - (\theta_1 + \theta_2)) - 2(x_3 - (\theta_1 + \theta_2)) \\ &\quad - 2(x_4 - \theta_1) - 2(x_5 - \theta_1) \\ &= -2\left[\sum_{i=1}^{5} x_i - 5\theta_1 - 3\theta_2 \right] \\ \frac{\partial}{\partial \theta_2} Q(\theta_1, \theta_2) &= -2(x_1 - (\theta_1 + \theta_2)) - 2(x_2 - (\theta_1 + \theta_2)) - 2(x_3 - (\theta_1 + \theta_2)) \\ &= -2\left[x_1 + x_2 + x_3 - 3\theta_1 - 3\theta_2 \right] \end{aligned}$$

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

$$\theta_1^* = \frac{x_4 + x_5}{2} \qquad \theta_2^* = \frac{x_1 + x_2 + x_3}{3} - \theta_1^*.$$

12.26 När man skall härleda ML- eller MK-skattningen, sätt aldrig in siffror direkt, utan först då slutresultat erhållits. Därför kalla vi observationerna $x_1, x_2, \ldots, x_7$. Minsta-kvadratmetoden: finn 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.

$$\begin{aligned} \frac{d}{d\theta}Q(\theta) &= -2\sum_{i=1}^{3}(x_i - \theta) - 2\sum_{i=4}^{7}(90 - x_i - \theta) \\ &= -2(x_1 + x_2 + x_3 + (90 - x_4) + \cdots + (90 - x_7) - 7\theta) = 0. \end{aligned}$$

Detta ger

$$\theta^* = \frac{x_1 + x_2 + x_3 + (90 - x_4) + \cdots + (90 - x_7)}{7},$$

eller, med siffror, $\theta^* = 61.17$. MK-skattningen av vinkeln vid C blir då $90 - 61.17 = 28.83$.

Skattningen beskrivs av skattningsvariabeln (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

$$\begin{aligned} E(\theta^*) &= E\left(\frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}\right) \\ &= \frac{1}{7}\big(\underbrace{E(X_1)}_{=\theta} + \cdots + \underbrace{E(X_3)}_{=\theta} + (90 - \underbrace{E(X_4)}_{=90-\theta}) + \cdots + (90 - \underbrace{E(X_7)}_{90-\theta})\big) = \frac{1}{7}\cdot 7\theta = \theta, \end{aligned}$$

vilket innebär att skattningen är väntevärdesriktig. Vidare,

$$\begin{aligned} V(\theta^*) &= V\left(\frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}\right) \\ &= \frac{1}{7^2}\big(\underbrace{V(X_1)}_{=\sigma^2} + \cdots + \underbrace{V(X_3)}_{=\sigma^2} + (-1)^2\underbrace{V(X_4)}_{=\sigma^2} + \cdots + (-1)^2\underbrace{V(X_7)}_{=\sigma^2}\big) = \frac{\sigma^2}{7}. \end{aligned}$$

Problemet är ett exempel på så kallade linjära modeller. Man kan visa att för linjära modeller är MK-skattningen den effektivaste, det vill säga den har minst varians av alla skattningar som kan skrivas som linjärkombinationer av data.

12.27 Låt X_1, X_2, X_3 beskriva mätningar av sträckan PQ och X_4 en mätning av sträckan QR. Alla mätningar sker utan systematiska fel. Modell: Alla stokastiska variabler $X_1, \ldots, X_4$ är oberoende, har samma varians och

$$E(X_1) = E(X_2) = E(X_3) = a \quad \text{och} \quad E(X_4) = a\cos(30^\circ) = \frac{a}{2}.$$

Minsta kvadratmetodens skattning av a är det värde på a som minimerar

$$Q(a) = \sum_{i=1}^{4}(x_i - E(X_i))^2 = (x_1 - a)^2 + (x_2 - a)^2 + (x_3 - a)^2 + (x_4 - a/2)^2.$$

Lösning av

$$0 = \frac{d}{da}Q(a) = -2(x_1 - a) - 2(x_2 - a) - 2(x_3 - a) - (x_4 - a/2) = \frac{-1}{2}\left[4(x_1 + x_2 + x_3) + 2x_4 - 13a\right]$$

ger

$$a = \frac{4(x_1 + x_2 + x_3) + 2x_4}{13}.$$

Med observationer:

$$x_1 = 452.0 \quad x_2 = 451.6 \quad x_3 = 451.7 \quad x_4 = 226.1$$

fås skattningen

$$a^* = \frac{4(x_1 + x_2 + x_3) + 2x_4}{13} = 451.8.$$

Punkten Q:s koordinater ges av $(x_Q, y_Q) = (\sqrt{3}a/2, a/2)$ som skattas med

$$(x_Q^*, y_Q^*) = (\sqrt{3}a^*/2, a^*/2) = (391.27, 225.90)$$

Skattningen a^* beskrivs av

$$A^* = \frac{4(X_1 + X_2 + X_3) + 2X_4}{13}$$

som har väntevärde

$$E(A^*) = E\left(\frac{4(X_1 + X_2 + X_3) + 2X_4}{13}\right) = \frac{4(a + a + a) + 2\,a/2}{13} = a,$$

dvs skattningen a^* är väntevärdesriktig, och varians

$$V(A^*) = V\left(\frac{4(X_1 + X_2 + X_3) + 2X_4}{13}\right) = \frac{1}{13^2}(4^2\sigma^2 + 4^2\sigma^2 + 4^2\sigma^2 + 2^2\sigma^2) = \frac{52}{13^2}\sigma^2 = \frac{4}{13}\sigma^2.$$

Alltså är

$$V\left(\frac{\sqrt{3}}{2}A^*\right) = \frac{3}{4}V(A^*) = \frac{3}{4}\,\frac{4}{13}\sigma^2 = \frac{3}{13}\sigma^2$$

och

$$V\left(\frac{1}{2}A^*\right) = \frac{1}{4}V(A^*) = \frac{1}{4}\,\frac{4}{13}\sigma^2 = \frac{1}{13}\sigma^2.$$

Skattningen av σ^2 ges av

$$s^2 = \frac{1}{4-1}\left(\sum_{i=1}^{3}(x_i - a^*)^2 + (x_4 - a^*/2)^2\right) = \frac{0.13}{3}.$$

Vi får tabellen:

koordinat	variansskattning	medelfel
x_Q	$3s^2/13 = 0.01$	0.1
y_Q	$s^2/13 = 0.0033333$	0.057735

12.28 Låt X_1, X_2 beskriva mätresultaten av sträckan PQ med observerade värden $x_1 = 52.1$ och $x_2 = 52.0$. Låt Y_1, Y_2 beskriva mätresultaten av sträckan RQ med observerade värden $y_1 = 60.8$ och $y_2 = 60.9$. Alla stokastiska variabler förutsätts vara oberoende, ha samma varians σ^2 och vara utan systematiska fel.

Med θ_x som sträckan PQ skattas θ_x med det värde som minimerar

$$Q(\theta_x) = \sum_{i=1}^{2}(x_i - E(X_i))^2 = (x_1 - \theta_x)^2 + (x_2 - \theta_x)^2.$$

Lösning av

$$0 = \frac{d}{d\theta_x}Q(\theta_x) = \sum_{i=1}^{2} 2(x_i - \theta_x)(-1) = (-2)((x_1 + x_2) - 2\theta_x)$$

ger $\theta_x = \overline{x}$. Teckenstudium av andraderivatan ger att detta verkligen är ett minimum. Alltså, skattningen av sträckan PQ är $\theta_x^* = \overline{x} = 52.05$. På samma sätt fås skattningen av sträckan QR, θ_y, som $\theta_y^* = \overline{y} = 60.85$.

Punkten Q har koordinaterna (x_q, y_q) som uttryckt i sträckan PQ är $(\theta_x/\sqrt{2}, \theta_x/\sqrt{2})$ och skattas med $(\theta_x^*/\sqrt{2}, \theta_x^*/\sqrt{2}) = (36.8, 36.8)$.

Punkten R har koordinaterna (x_r, y_r) som uttryckt i sträckan PQ och sträckan QR är $(\theta_x/\sqrt{2}+\sqrt{3}\theta_y/2, \theta_x/\sqrt{2}+\theta_y/2)$ och skattas med $(\theta_x^*/\sqrt{2} + \sqrt{3}\theta_y^*/2, \theta_x^*/\sqrt{2} + \theta_y^*/2) = (89.5, 67.2)$.

$$\begin{aligned}
V\left(x_q^*\right) &= V\left(\theta_x^*/\sqrt{2}\right) = \frac{1}{2}V\left(\theta_x^*\right) = \frac{1}{2}V\left(\frac{1}{2}(X_1 + X_2)\right) = \frac{1}{8}V\left(X_1 + X_2\right) = \frac{1}{8}\sigma^2 + \frac{1}{8}\sigma^2 \\
&= \frac{1}{4}\sigma^2. \\
V\left(y_q^*\right) &= V\left(\theta_x^*/\sqrt{2}\right) = V\left(x_q^*\right) = \frac{1}{4}\sigma^2. \\
V\left(x_r^*\right) &= V\left(\theta_x^*/\sqrt{2} + \sqrt{3}/2\theta_y^*\right) = \frac{1}{2}V\left(\theta_x^*\right) + \frac{3}{4}V\left(\theta_y^*\right) \\
&= \frac{1}{2}V\left(\frac{1}{2}(X_1 + X_2)\right) + \frac{3}{4}V\left(\frac{1}{2}(Y_1 + Y_2)\right) = \sigma^2/4 + 3\sigma^2/8 = \frac{5}{8}\sigma^2 \\
V\left(y_r^*\right) &= V\left(\theta_x^*/\sqrt{2} + \theta_y^*/2\right) = \frac{1}{2}V\left(\theta_x^*\right) + \frac{1}{4}V\left(\theta_y^*\right) \\
&= \frac{1}{2}V\left(\frac{1}{2}(X_1 + X_2)\right) + \frac{1}{4}V\left(\frac{1}{2}(Y_1 + Y_2)\right) = \frac{\sigma^2}{4} + \frac{\sigma^2}{8} = \frac{3}{8}\sigma^2.
\end{aligned}$$

Skattningsvariabeln av σ^2 är

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + (Y_1 - \overline{Y})^2 + (Y_2 - \overline{Y})^2}{2}$$

med observerat värde $s^2 = 0.005$. Vi får medelfelen i koordinatskattningarna

koordinat	variansskattning	medelfel
x_q	$s^2/4 = 0.00125$	0.035355
y_q	$s^2/4 = 0.00125$	0.035355
x_r	$5s^2/8 = 0.003125$	0.055902
y_r	$3s^2/8 = 0.001875$	0.043301

12.29 Låt $x_1, \ldots, x_n$ vara observationerna, utfall av oberoende normalfördelade stokastiska variabler $X_1, \ldots, X_n$. Normalfördelningens täthetsfunktion är $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m)^2/2\sigma^2}$. Likelihoodfunktionen blir därför

$$\begin{aligned}
L(m, \sigma) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \{\text{oberoende}\} = f_{X_1}(x_1)\cdots f_{X_n}(x_n) \\
&= \frac{1}{\sqrt{2\pi}\sigma}e^{-(x_1-m)^2/2\sigma^2}\cdots\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_n-m)^2/2\sigma^2} = \frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2}
\end{aligned}$$

Logaritmering ger

$$\ln(L(m,\sigma)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2.$$

För att bestämma maximum deriveras denna med avseende på m och σ.

$$\frac{\partial}{\partial m}\ln(L(m,\sigma)) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m) = \frac{n}{\sigma^2}\left[\overline{x} - m\right].$$

Om denna sätts till 0 fås skattningen $m^*(x_1,\ldots,x_n) = \overline{x}$ av m. Om man utnyttjar denna skattning ger derivering med avseende på σ sedan

$$\frac{\partial}{\partial \sigma}\ln(L(\overline{x},\sigma)) = -n\frac{1}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{-n}{\sigma^3}\left[\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right],$$

dvs om derivatan sätts till 0 fås skattningen

$$\sigma^*(x_1,\ldots,x_n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

av σ.

Sätter vi in siffror erhåller vi $m^* = 250.32$ och $\sigma^* = 2.124$.

12.30 Låt $X_1,\ldots,X_n$ beskriva mätresultaten för proverna från det första partiet och $Y_1,\ldots,Y_r$ motsvarande för det andra partiet, med observerade värden $x_1,\ldots,x_n$ resp. $y_1,\ldots,y_r$. Modell: Alla stokastiska variabler är oberoende och

$$E\left(X_i\right) = m_x \quad V\left(X_i\right) = \sigma_x^2 \qquad E\left(Y_i\right) = m_y \quad V\left(Y_i\right) = \sigma_y^2.$$

En skattning av m_x är $\overline{x}$ och en skattning av m_y är $\overline{y}$. En skattning av $m_x - m_y$ ges av $\overline{x} - \overline{y}$ som beskrivs av $\overline{X} - \overline{Y}$ som har

$$\begin{aligned} E\left(\overline{X} - \overline{Y}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{r}\sum_{j=1}^{r}Y_j\right) = \frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right) - \frac{1}{r}\sum_{j=1}^{r}E\left(Y_j\right) = \frac{1}{n}\sum_{i=1}^{n}m_x - \frac{1}{r}\sum_{j=1}^{r}m_y \\ &= m_x - m_y \end{aligned}$$

och

$$\begin{aligned} V\left(\overline{X} - \overline{Y}\right) &= V\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{r}\sum_{j=1}^{r}Y_j\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n}V\left(X_i\right) + \frac{(-1)^2}{r^2}\sum_{j=1}^{r}V\left(Y_j\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_x^2 + \frac{1}{r^2}\sum_{j=1}^{r}\sigma_y^2 = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{r}. \end{aligned}$$

Med data:

$$\overline{x} = 24.2 \quad s_x = 1.6 \qquad \overline{y} = 28.5 \quad s_y = 2.3$$

fås skattningen av $m_x - m_y$ till

$$\overline{x} - \overline{y} = 24.2 - 28.5 = -4.30.$$

Standardavvikelsen $D\left(\overline{X} - \overline{Y}\right) = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{r}}$ skattas med

$$\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{r}} = \sqrt{\frac{1.6^2}{5} + \frac{2.3^2}{10}} = 1.0203$$

vilket är medelfelet för skattningen $\overline{x} - \overline{y}$.

På samma sätt: En skattning av $(m_x+m_y)/2$ ges av $(\overline{x}+\overline{y})/2$ som beskrivs av $(\overline{X}+\overline{Y})/2$ som har

$$\begin{aligned} E\left(\frac{\overline{X}+\overline{Y}}{2}\right) &= E\left(\frac{1}{2n}\sum_{i=1}^{n}X_i+\frac{1}{2r}\sum_{j=1}^{r}Y_j\right)=\frac{1}{2n}\sum_{i=1}^{n}E\left(X_i\right)+\frac{1}{2r}\sum_{j=1}^{r}E\left(Y_j\right) \\ &= \frac{1}{2n}\sum_{i=1}^{n}m_x+\frac{1}{2r}\sum_{j=1}^{r}m_y=\frac{m_x+m_y}{2}. \end{aligned}$$

och

$$\begin{aligned} V\left(\frac{\overline{X}+\overline{Y}}{2}\right) &= V\left(\frac{1}{2n}\sum_{i=1}^{n}X_i+\frac{1}{2r}\sum_{j=1}^{r}Y_j\right)=\{\text{oberoende}\} \\ &= \frac{1}{(2n)^2}\sum_{i=1}^{n}V\left(X_i\right)+\frac{1}{(2r)^2}\sum_{j=1}^{r}V\left(Y_j\right)=\frac{1}{4n^2}\sum_{i=1}^{n}\sigma_x^2+\frac{1}{4r^2}\sum_{j=1}^{r}\sigma_y^2 \\ &= \frac{\sigma_x^2}{4n}+\frac{\sigma_y^2}{4r}. \end{aligned}$$

Med data fås skattningen av $(m_x+m_y)/2$ till

$$\frac{\overline{x}+\overline{y}}{2}=\frac{24.2+28.5}{2}=26.35$$

Standardavvikelsen $D\left((\overline{X}+\overline{Y})/2\right)=\sqrt{\frac{\sigma_x^2}{4n}+\frac{\sigma_y^2}{4r}}$ skattas med

$$\sqrt{\frac{s_x^2}{4n}+\frac{s_y^2}{4r}}=\sqrt{\frac{1.6^2}{4\cdot 5}+\frac{2.3^2}{4\cdot 10}}=0.51015$$

vilket är medelfelet för skattningen $(\overline{x}+\overline{y})/2$.

12.31 Observationer på legering i beskrivs av n_i stycken oberoende normalfördelade stokastiska variabler med väntevärde m_i och standardavvikelse σ. Observerade värden sammanfattas i

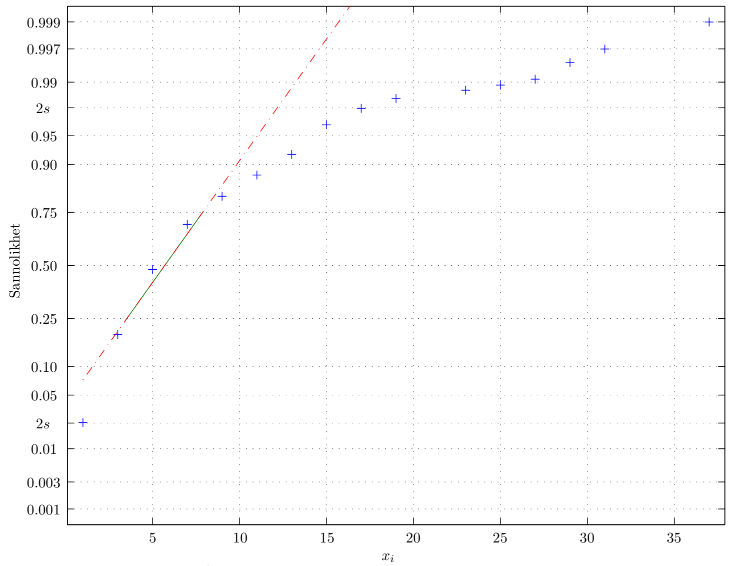
Legering i	Antal obs n_i	Medelvärde $\overline{x}_i$	Stickprovsvarians s_i^2	s_i
1	3	128.83	0.013333	0.11547
2	3	135.50	0.010000	0.10000
3	3	140.70	0.030000	0.17321
4	3	137.07	0.0033333	0.057735

där medelvärdena $\overline{x}_i$ är skattningar av smältpunkterna m_i. Variansskattningarna poolas samman till en skattning av σ^2:

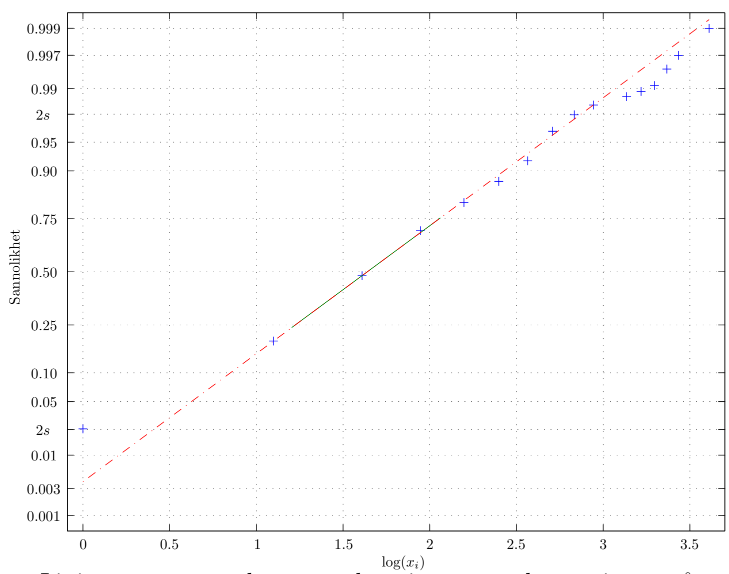
$$s^2=\frac{(n_1-1)s_1^2+\cdots+(n_4-1)s_4^2}{(n_1-1)+\cdots+(n_4-1)}=\frac{3s_1^2+\cdots+3s_4^2}{12}=0.014167$$

vilket ger $s=0.11902$.

12.32 De 497 klassindelade observationerna plottas i följande figurer.



Plot av data på normalfördelningspapper. Linjen motsvarar data från normalfördelning. Data följer inte linjen och modelleras dåligt av en normalfördelning.



Linjen motsvarar data vars logaritmer är observationer på en normalfördelning. Data följer linjen bra och lognormalfördelningen kan vara en lämplig modell för data.

13.1 Låt X vara $\mathrm{N}(0,1)$. Bestäm a så att $P\left(|X| \leq a\right) = 0.95$. Då är

$$0.95 \quad = \quad P\left(|X| \leq a\right) = P\left(-a \leq X \leq a\right) = \Phi\left(a\right) - \Phi\left(-a\right) = \Phi\left(a\right) - \left(1 - \Phi\left(a\right)\right) = 2\Phi\left(a\right) - 1.$$

Alltså är $\Phi\left(a\right) = 0.975$ och

$$0.025 = 1 - \Phi\left(a\right) = 1 - \Phi\left(\lambda_{0.025}\right)$$

dvs $a = \lambda_{0.025} = 1.9600$.

Med

$$0.99 \quad = \quad P\left(|X| \leq b\right) = P\left(-b \leq X \leq b\right) = \Phi\left(b\right) - \Phi\left(-b\right) = \Phi\left(b\right) - \left(1 - \Phi\left(b\right)\right) = 2\Phi\left(b\right) - 1.$$

Alltså är $\Phi(b) = 0.995$ och

$$0.005 = 1 - \Phi(b) = 1 - \Phi(\lambda_{0.005})$$

dvs $b = \lambda_{0.005} = 2.5758$.

13.2 Låt X vara $\chi^2(24)$-fördelad. Bestäm a så att $P(X < a) = 0.95$. Då är

$$0.05 = 1 - P(X \leq a) = 1 - F_X(a) = 1 - F_X(\chi^2_{0.05})$$

dvs $a = \chi^2_{0.05} = 36.415$.

Bestäm b och 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

$$0.025 = 1 - P(X \leq c) = 1 - F_X(c) = 1 - F_X(\chi^2_{0.025})$$

dvs $c = \chi^2_{0.025} = 39.364$ och

$$0.975 = 1 - P(X \leq b) = 1 - F_X(b) = 1 - F_X(\chi^2_{0.975})$$

dvs $b = \chi^2_{0.975} = 12.401$.

13.3 Om $Z, Z_1, Z_2, \ldots, Z_n$ är oberoende $\mathrm{N}(0, 1)$ så är

$$X = \sum_{i=1}^{n} Z_i^2$$

$\chi^2(n)$-fördelad. För Z är

$$E\left(Z^2\right) = V(Z) + (E(Z))^2 = 1 + 0^2 = 1$$

så

$$E(X) = E\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} \underbrace{E\left(Z_i^2\right)}_{=1} = n.$$

Vidare så är

$$\begin{aligned} E\left(Z^4\right) &= \int z^4 f_Z(z)\,dz = \int_{-\infty}^{\infty} z^4 \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\,dz = \{u = z^2/2\} \\ &= 2 \cdot 2^{3/2} \frac{1}{\sqrt{2\pi}} \underbrace{\int_0^{\infty} u^{3/2}\,\mathrm{e}^{-u}\,du}_{=\Gamma(5/2)} = \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right). \end{aligned}$$

Nu är

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\,\frac{1}{2}\,\underbrace{\Gamma\left(\frac{1}{2}\right)}_{=\sqrt{\pi}} = \frac{3}{4}\sqrt{\pi}$$

så $E\left(Z^4\right) = 3$ och

$$V\left(Z^2\right) = E\left(Z^4\right) - \left(E\left(Z^2\right)\right)^2 = 3 - 1^2 = 2$$

och

$$V(X) = V\left(\sum_{i=1}^{n} Z_i^2\right) = \{\text{oberoende}\} = \sum_{i=1}^{n} \underbrace{V\left(Z_i^2\right)}_{=2} = 2n.$$

13.4 Om $Z, Z_1, Z_2, \ldots, Z_n$ är oberoende $\mathrm{N}(0, 1)$ så är

$$X = \sum_{i=1}^{n} Z_i^2$$

$\chi^2(n)$-fördelad. Ur uppgift 13.3 är

$$E(X) = n \qquad V(X) = 2n,$$

och enligt centrala gränsvärdessatsen är X approximativt normalfördelad, X är approx. $\mathrm{N}(n, \sqrt{2n})$. Det vill säga

$$1 - \Phi(\lambda_\alpha) = \alpha = P\left(X > \chi^2_\alpha\right) = P\left(\frac{X-n}{\sqrt{2n}} > \frac{\chi^2_\alpha - n}{\sqrt{2n}}\right) \approx 1 - \Phi\left(\frac{\chi^2_\alpha - n}{\sqrt{2n}}\right)$$

ger att

$$\lambda_\alpha = \frac{\chi^2_\alpha - n}{\sqrt{2n}}$$

eller

$$\chi^2_\alpha = n + \sqrt{2n}\lambda_\alpha.$$

13.5 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
1	2.0	28
2	1.96	473
3	1.960	4427
4	1.9600	27581

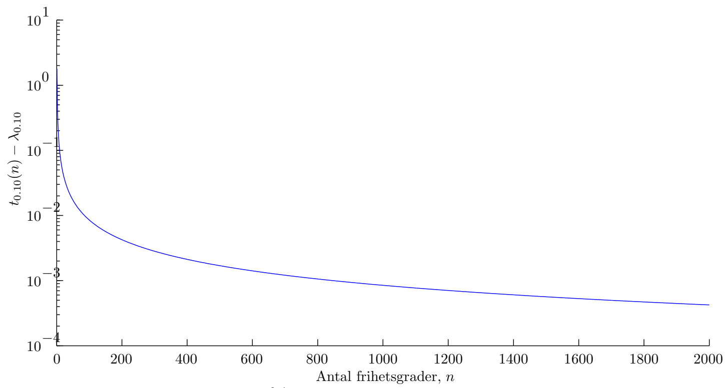
Kvantil $\alpha = 0.05$

# decimaler	$\lambda_\alpha = t_\alpha$	# frihetsgrader
1	1.6	298
2	1.64	10412
3	1.645	10412
4	1.6449	15813

Kvantil $\alpha = 0.10$

# decimaler	$\lambda_\alpha = t_\alpha$	# frihetsgrader
1	1.3	14
2	1.28	247
3	1.282	894
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 mellan 10%-kvantilerna i t-fördelningen och normalfördelningen för växande antal frihetsgrader.

13.6 Låt X beskriva resultatet av en avståndsmätning. Modell:

$$X = \text{avstånd} + \text{mätfel} = m + \epsilon$$

där ϵ är $\mathrm{N}(0,\sigma)$, $\sigma = 5 \cdot 10^{-3}$. Då är X $\mathrm{N}(m,\sigma)$. Låt $X_1,\ldots,X_n$ beskriva resultaten av n oberoende avståndsmätningar med observerade värden $x_1,\ldots,x_n$. Väntevärdet $m = E(X)$ skattas med $\overline{x}$ som beskrivs av $\overline{X}$, där

$$\overline{X} \quad \text{är} \quad \mathrm{N}\left(m, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{\overline{X} - m}{\sigma/\sqrt{n}} \quad \mathrm{N}(0,1)$$

så med sannolikhet $1-\alpha$ är

$$-\lambda_{\alpha/2} \le \frac{\overline{X} - m}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1-\alpha$ är

$$\overline{X} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le m \le \overline{X} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\overline{x} = 1132.155$. Konfidensgrad $1 - \alpha = 0.95$ ger $\lambda_{\alpha/2} = \lambda_{0.025} = 1.9600$ och intervallet

$$m \in \overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1132.155 \pm 1.96\,\frac{5 \cdot 10^{-3}}{\sqrt{4}} = 1132.155 \pm 0.0048999 \qquad (95\%)$$

eller

$$1132.150 \le m \le 1132.160 \qquad (95\%).$$

13.8 Låt $X_1,\ldots,X_n$ beskriva mätresultaten med observerade värden $x_1,\ldots,x_n$, $n = 6$. Modell: X_i är oberoende och $\mathrm{N}(m,\sigma)$ där $\sigma = 0.2$. Alltså är $\overline{X}$ $\mathrm{N}(m, \sigma/\sqrt{n})$ och

$$\frac{\overline{X} - m}{\sigma/\sqrt{n}} \text{ är } \mathrm{N}(0,1).$$

Således, med sannolikhet $1-\alpha$ är

$$-\lambda_{\alpha/2} \le \frac{\overline{X} - m}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2},$$

vilket omskrivet ger att med sannolikhet $1-\alpha$ är

$$\overline{X} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le m \le \overline{X} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Med $1-\alpha = 0.90$ fås $\lambda_{\alpha/2} = \lambda_{0.05} = 1.6449$ så med sannolikhet 90% är

$$m \in \overline{X} \pm 1.6449 \cdot \frac{0.2}{\sqrt{6}} = \overline{X} \pm 0.1343.$$

Med observerat värde $\overline{x} = 5.575$ fås intervallet

$$m \in 5.575 \pm 0.1343 = [5.4407,\ 5.7093] \qquad (90\%).$$

Vi fortsatta dagar räcker det med att beräkna $\overline{x}$.

13.9 Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(m, \sigma)$, där σ är känd, med observerade värden $x_1, \ldots, x_n$.

a) Ett konfidensintervall för m med konfindensintervall $1-\alpha$ ges av

$$m \in \overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Konfidensintervallet har bredden

$$L = 2 \cdot \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Antalet observationer n som motsvarar en given bredd L ges av

$$n = \left(2 \cdot \lambda_{\alpha/2}\frac{\sigma}{L}\right)^2 = 4\lambda_{\alpha/2}^2\frac{\sigma^2}{L^2}.$$

b) Med konfidensgrad $1-\alpha = 0.95$ och $\sigma = 1$ fås $\lambda_{0.025} = 1.96$ och

$$n = 4 \cdot (1.96)^2 \cdot \frac{1^2}{L^2} = \frac{15.3664}{L^2}.$$

Med insatta värden på L:

L	n
1	16
0.5	62
0.1	1537

(avrunda n uppåt till heltal)

13.10 Konfidensintervallet för väntevärdet med konfidensgrad $1-\alpha$ ges av

$$\overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

dvs 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 $L(0.90, 5)/10$].

$$2\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2\lambda_{\alpha/2}\frac{\sigma}{\sqrt{5}}\Big/ 2$$

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 $\lambda_{0.005} = 2.5758$, $n = 5(\lambda_{0.005}/\lambda_{0.05})^2 = 12.26$, dvs. 13 stycken.

c) Vi söker n så att $L(0.99, n) = L(0.90, 5)/2$ [alternativt $L(0.90, 5)/10$].

$$2\lambda_{0.005}\frac{\sigma}{\sqrt{n}} = 2\lambda_{0.05}\frac{\sigma}{\sqrt{5}}\Big/ 2$$

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 $\overline{x}$ eller σ.

13.11 Låt $X_1, \ldots, X_n$ beskriva avkastningen hos den kemiska industrin med observerade värden $x_1, \ldots, x_n$. Modell: X_i är oberoende och $\mathrm{N}(m, \sigma)$. Alltså är $\overline{X}$ $\mathrm{N}(m, \sigma/\sqrt{n})$ och

$$\frac{\overline{X} - m}{S/\sqrt{n}} \text{ är } t(n-1)$$

där S^2 är stickprovsvariansen. Således, med sannolikhet $1 - \alpha$ är

$$-t_{\alpha/2} \le \frac{\overline{X} - m}{S/\sqrt{n}} \le t_{\alpha/2},$$

vilket omskrivet ger att med sannolikhet $1 - \alpha$ är

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \le m \le \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}.$$

Med $n = 10$ observationer och konfidensgrad $1 - \alpha = 0.95$ fås ur $t(n-1) = t(9)$-tabell att $t_{\alpha/2} = \lambda_{0.025} = 2.26$. Vidare beräknas

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 7.51 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0.1543$$

$s = \sqrt{s^2} = 0.3929$, så det observerade intervallet blir

$$m \in \overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 7.51 \pm 0.281 = [7.229,\ 7.791] \qquad (95\%).$$

13.12 Låt $x_1, \ldots, x_n$ vara utfall på $X_1, \ldots, X_n$ där X_i beskriver längden av planka i. Modell: $X_1, \ldots, X_n$ är oberoende och $\mathrm{N}(m, \sigma)$. Vi skattar m med $\overline{x}$ som beskrivs av $\overline{X}$ som är $\mathrm{N}(m, \sigma/\sqrt{n})$. Alltså är

$$\frac{\overline{X} - m}{\sigma/\sqrt{n}} \quad \mathrm{N}(0, 1)$$

och

$$\frac{\overline{X} - m}{S/\sqrt{n}} \quad t(n-1)\text{-fördelad.}$$

Alltså är med sannolikhet $1 - \alpha$

$$-t_{\alpha/2}(n-1) \le \frac{\overline{X} - m}{S/\sqrt{n}} \le t_{\alpha/2}(n-1)$$

eller

$$m \in \overline{X} \pm t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}.$$

Med observationer $x_1, \ldots, x_n$, $n = 16$:

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

fås skattningarna

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 4.6812$$

och

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{0.46962} = 0.68529.$$

Med konfidensgrad $1-\alpha = 0.95$ fås att $t_{\alpha/2}(15) = t_{0.025}(15) = 2.1314$ och konfidensintervallet blir

$$m \in 4.6812 \pm 2.1314\,\frac{0.68529}{\sqrt{16}} = 4.6812 \pm 0.36517 = [4.3161,\ 5.0464] \quad (95\%).$$

13.13 Låt X beskriva resultatet av en fryspunktsmätning. Modell:

$$X = \text{fryspunkt} + \text{mätfel} = m + \epsilon$$

där ϵ är $\mathrm{N}(0,\sigma)$. Då är X $\mathrm{N}(m,\sigma)$. Låt $X_1, \ldots, X_n$ beskriva resultaten av n oberoende fryspunktsmätningar med observerade värden $x_1, \ldots, x_n$. Väntevärdet $m = E\,(X)$ skattas med $\overline{x}$ som beskrivs av $\overline{X}$, där

$$\overline{X} \quad \text{är} \quad \mathrm{N}\left(m, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{\overline{X}-m}{\sigma/\sqrt{n}} \quad \mathrm{N}(0,1)$$

så med sannolikhet $1-\alpha$ är

$$-\lambda_{\alpha/2} \le \frac{\overline{X}-m}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1-\alpha$ är

$$\overline{X} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le m \le \overline{X} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Med $n = 5$ observationer fås $\overline{x} = 1.3$. Konfidensgrad $1-\alpha = 0.90$ ger $\lambda_{\alpha/2} = \lambda_{0.05} = 1.6445$ och intervallet

$$m \in \overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.3 \pm 1.6445\,\frac{\sqrt{0.7}}{\sqrt{5}} = 1.3 \pm 0.615 \qquad (90\%)$$

eller

$$0.685 \le m \le 1.92 \qquad (90\%).$$

Om σ^2 är okänd och skattas med

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) = 0.7$$

utnyttjas att

$$\frac{\overline{X} - m}{S/\sqrt{n}} \quad \text{är} \quad t(n-1)$$

så med sannolikhet $1 - \alpha$ är

$$-t_{\alpha/2}(n-1) \leq \frac{\overline{X} - m}{S/\sqrt{n}} \leq t_{\alpha/2}(n-1)$$

eller, omformat, med sannolikhet $1 - \alpha$ är

$$\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \leq m \leq \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}.$$

Konfidensgrad $1 - \alpha = 0.90$ ger $t_{\alpha/2}(n-1) = t_{0.05}(4) = 2.13$ och intervallet

$$m \in \overline{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = 1.3 \pm 2.13\,\frac{\sqrt{0.7}}{\sqrt{5}} = 1.3 \pm 0.798 \qquad (90\%)$$

eller

$$0.502 \leq m \leq 2.10 \qquad (90\%).$$

Med $n = 25$ mätningar och $\sigma^2 = 0.7$ fås intervallet

$$m \in \overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.3 \pm 1.6445\,\frac{\sqrt{0.7}}{\sqrt{25}} = 1.3 \pm 0.275 \qquad (90\%).$$

Om σ^2 är okänd och skattas med $s^2 = 0.7$ fås, med $t_{0.05}(24) = 1.71$, att

$$m \in \overline{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = 1.3 \pm 1.71\,\frac{\sqrt{0.7}}{\sqrt{25}} = 1.3 \pm 0.286 \qquad (90\%)$$

13.14 Låt X_{ij} beskriva resultatet av den j:te upprepningen av mätning i. Modell:

$$X_{ij} = \text{fysikalisk konstant} + \text{mätfel} = m + \epsilon_{ij}$$

där ϵ_{ij} är oberoende $\mathrm{N}(0, \sigma)$, dvs X_{ij} är oberoende $\mathrm{N}(m, \sigma)$. Observationer x_{ij} på X_{ij} ges av tabellen

	Mätning i: 1	2	3	4	5
j = 1	24.3	20.8	17.5	19.0	20.7
2	23.4	20.1	16.5	19.4	19.3
3	23.4	20.9	17.6	19.2	19.7
$y_i = \overline{x}_i$	23.7	20.6	17.2	19.2	19.9

Med

$$y_i = \frac{1}{n_j}\sum_{j=1}^{n_j} x_{ij} = \overline{x}_i, \quad i = 1, \ldots, 5$$

fås $k = 5$ stycken oberoende skattningar av m. Vidare har vi att $\overline{y} = 20.12$ och $s_y = 2.37$ så med konfidensgrad $1 - \alpha = 0.95$ fås $t_{\alpha/2}(k-1) = t_{0.025}(4) = 2.78$ och konfidensintervallet

$$m \in \overline{y} \pm t_{0.025}(4)\frac{s_y}{\sqrt{5}} = m \in 20.1 \pm 2.78\frac{2.37}{\sqrt{5}} = 20.1 \pm 2.94 \qquad (95\%)$$

eller $17.1 \leq m \leq 23.1$ (95%).

13.15 Vi vet att

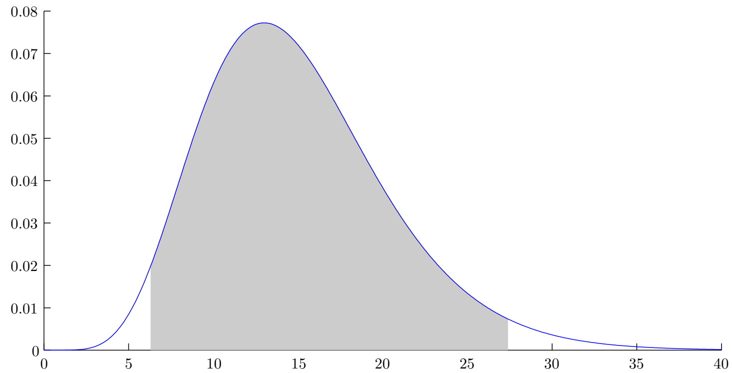
$$\frac{(n-1)S^2}{\sigma^2} \quad \text{är} \quad \chi^2(n-1)\text{-fördelad}$$

om $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ där X_i är oberoende $\mathrm{N}(m, \sigma)$. Alltså är

$$P\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Med $\alpha = 0.05$ fås ur $\chi^2(n-1) = \chi^2(15)$-tabeller

$$P\left(6.26 \le \frac{(n-1)S^2}{\sigma^2} \le 27.5\right) = 0.95.$$



Kvantiler i $\chi^2(15)$-fördelningen. På intervallet $(6.26,\ 27.5)$ ligger 95% av fördelningens sannolikhetsmassa.

Alltså med sannolikhet 95% är

$$\frac{(n-1)S^2}{27.5} \le \sigma^2 \le \frac{(n-1)S^2}{6.26}.$$

Med skattningen $s^2 = 0.46962$ fås intervallet

$$0.25627 = \frac{(n-1)s^2}{27.5} \le \sigma^2 \le \frac{(n-1)s^2}{6.26} = 1.1249 \qquad (95\%)$$

eller

$$0.50623 = \sqrt{0.25627} \le \sigma \le \sqrt{1.1249} = 1.0606 \qquad (95\%).$$

13.16 Vi vet att

$$\frac{(n-1)S^2}{\sigma^2} \quad \text{är} \quad \chi^2(n-1)\text{-fördelad}$$

om $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ där X_i är oberoende $\mathrm{N}(m, \sigma)$. Alltså är

$$P\left(\frac{(n-1)S^2}{\sigma^2} \ge \chi^2_{1-\alpha}\right) = 1 - \alpha.$$

Med $\alpha = 0.95$ fås ur $\chi^2(n-1) = \chi^2(49)$-tabeller

$$P\left(\frac{(n-1)S^2}{\sigma^2} \ge 33.93\right) = 0.95.$$

Alltså med sannolikhet 95% är

$$\sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha}} \qquad \text{dvs} \qquad \sigma^2 \le \frac{(n-1)S^2}{33.93}.$$

Med skattningen $s = 0.021$ fås intervallet

$$\sigma^2 \leq \frac{(n-1)s^2}{33.93} = 0.00063686 \qquad (95\%)$$

eller

$$\sigma \leq \sqrt{0.00063686} = 0.025236 \qquad (95\%).$$

13.17 Låt $X_1, \ldots, X_n$ beskriva mätresultaten med observerade värden $x_1, \ldots, x_n$. Modell: X_i är oberoende och $\mathrm{N}(\delta_1, \sigma_1)$. Alltså är $\overline{X}$ $\mathrm{N}(\delta_1, \sigma_1/\sqrt{n})$ och

$$\frac{\overline{X} - \delta_1}{S_x/\sqrt{n}} \text{ är } t(n-1)$$

där S_x^2 är stickprovsvariansen. Således, med sannolikhet $1-\alpha$ är

$$-t_{\alpha/2} \leq \frac{\overline{X} - \delta_1}{S_x/\sqrt{n}} \leq t_{\alpha/2},$$

vilket omskrivet ger att med sannolikhet $1-\alpha$ är

$$\overline{X} - t_{\alpha/2}\frac{S_x}{\sqrt{n}} \leq \delta_1 \leq \overline{X} + t_{\alpha/2}\frac{S_x}{\sqrt{n}}.$$

Med $n = 100$ observationer och konfidensgrad $1-\alpha = 0.99$ fås ur $t(n-1) = t(99)$-tabell (använd normalfördelningens tabell om t-tabellen inte räcker till) att $t_{\alpha/2} = t_{0.005} = 2.6264$ (2.5758 med normalfördelning) så det observerade intervallet blir

$$\delta_1 \in \overline{x} \pm t_{\alpha/2}\frac{s_x}{\sqrt{n}} = -0.28 \pm 0.5515 = [-0.8315,\ 0.2715] \qquad (99\%).$$

Motsvarande för δ_2 ges av

$$\delta_2 \in \overline{y} \pm t_{\alpha/2}\frac{s_y}{\sqrt{n}} = 0.11 \pm 0.998 = [-0.8880,\ 1.1080] \qquad (99\%).$$

Eftersom $(n-1)S_x^2/\sigma_1^2$ är $\chi^2(n-1)$-fördelad så är med sannolikhet $1-\alpha$

$$\chi^2_{1-\alpha/2} \leq \frac{(n-1)S_x^2}{\sigma_1^2} \leq \chi^2_{\alpha/2}$$

vilket omformat ger att med sannolikhet $1-\alpha$ är

$$\frac{(n-1)S_x^2}{\chi^2_{\alpha/2}} \leq \sigma_1^2 \leq \frac{(n-1)S_x^2}{\chi^2_{1-\alpha/2}}.$$

eller

$$\sqrt{\frac{(n-1)S_x^2}{\chi^2_{\alpha/2}}} \leq \sigma_1 \leq \sqrt{\frac{(n-1)S_x^2}{\chi^2_{1-\alpha/2}}}.$$

Med $n = 100$ observationer och konfidensgrad $1-\alpha = 0.99$ fås ur $\chi^2(n-1)$-tabell $\chi^2_{1-\alpha/2} = \chi^2_{0.995} = 66.51$ och $\chi^2_{\alpha/2} = \chi^2_{0.005} = 138.99$. (Om man inte har $\chi^2(n-1) = \chi^2(99)$-tabeller kan man utnyttja att för stora ν är kvantilerna $\chi^2_\alpha(\nu) \approx \nu + \lambda_\alpha\sqrt{2\nu}$ (se uppgift 13.4) vilket ger $\chi^2_{0.005} \approx 135.25$ och $\chi^2_{0.995} \approx 62.755$.)

Det observerade intervallet blir

$$1.7724 = \sqrt{\frac{(n-1)s_x^2}{\chi^2_{\alpha/2}}} \leq \sigma_1 \leq \sqrt{\frac{(n-1)s_x^2}{\chi^2_{1-\alpha/2}}} = 2.5621 \qquad (99\%).$$

Motsvarande för σ_2 är

$$3.2071 = \sqrt{\frac{(n-1)s_y^2}{\chi^2_{\alpha/2}}} \le \sigma_1 \le \sqrt{\frac{(n-1)s_y^2}{\chi^2_{1-\alpha/2}}} = 4.6361 \qquad (99\%).$$

Instrumentet som ligger bakom observationerna $y_1, \ldots, y_n$ verkar ha en större spridning, mindre noggrannhet. Båda instrumenten kan vara utan systematiska fel.

	största systematiska fel	standardavvikelse för slumpmässiga fel	max.fel
	$\lvert\delta\rvert$	$\lvert\sigma\rvert$	$\lvert\delta\rvert + 3\lvert\sigma\rvert$
instrument A	0.8315	2.5621	$0.8315 + 3 \cdot 2.5621 = 8.5178$
instrument B	1.1080	4.6361	$1.1080 + 3 \cdot 4.6361 = 15.0163$

13.18 Om $Z, Z_1, Z_2, \ldots, Z_n$ är oberoende N$(0,1)$ så är

$$X = \sum_{i=1}^{n} Z_i^2$$

$\chi^2(n)$-fördelad. Om X och X' är oberoende och $\chi^2(n)$- respektive $\chi^2(n')$-fördelade så är

$$X + X' = \sum_{i=1}^{n} Z_i^2 + \sum_{i=1}^{n'} Z_i'^2$$

en summa av $n + n'$ oberoende $(\mathrm{N}(0,1))^2$ termer och alltså är $\chi^2(n+n')$-fördelad. Eftersom

$$\frac{(n_i-1)S_i^2}{\sigma^2}$$

är $\chi^2(n_i - 1)$-fördelad så är

$$\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} + \cdots + \frac{(n_k-1)S_k^2}{\sigma^2} = \frac{(n_1-1)S_1^2 + \cdots + (n_k-1)S_k^2}{\sigma^2}$$

$\chi^2(n_1 + n_2 + \cdots + n_k - k)$-fördelad. Det vill säga, med $f = n_1 + n_2 + \cdots + n_k - k$ så är

$$S^2 = \frac{(n_1-1)S_1^2 + \cdots + (n_k-1)S_k^2}{f}$$

sådan att $f \cdot S^2/\sigma^2$ är $\chi^2(f)$-fördelad.

Alltså är

$$P\left(\chi^2_{1-\alpha/2} \le \frac{f\,S^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Alltså med sannolikhet $1 - \alpha$ är

$$\frac{f\,S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{f\,S^2}{\chi^2_{1-\alpha/2}}$$

eller

$$\underbrace{\sqrt{\frac{f}{\chi^2_{\alpha/2}}}}_{=k_1} S \le \sigma \le \underbrace{\sqrt{\frac{f}{\chi^2_{1-\alpha/2}}}}_{=k_2} S$$

eller $k_1 S \le \sigma \le k_2 S$.

13.19 Fem mätserier ger upphov till $k = 5$ oberoende skattningar $s_1, \dots, s_5$ av samma standardavvikelse σ.

Mätserie, i	1	2	3	4	5
Antal observationer, n_i	4	4	2	2	2
Stickprovsstandardavvikelse, s_i	0.11	0.15	0.09	0.20	0.07
Stickprovsvarians, s_i^2	0.0121	0.0225	0.0081	0.0400	0.0049

Som den sammanpoolade stickprovsvariansen fås

$$s^2 = \frac{(n_1 - 1)s_1^2 + \cdots + (n_5 - 1)s_5^2}{n_1 + \cdots + n_5 - 5} = \frac{0.052267 + \cdots + 0.017422}{9} = 0.017422,$$

dvs $s = 0.13199$.

Vi vet att

$$\frac{9\,S^2}{\sigma^2} \quad \text{är} \quad \chi^2(9)\text{-fördelad}$$

där S^2 är stickprovsvariabeln motsvarande s^2. Alltså är

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{9\,S^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Med $\alpha = 0.99$ fås ur $\chi^2(9)$-tabeller

$$P\left(1.73 \leq \frac{9\,S^2}{\sigma^2} \leq 23.6\right) = 0.99.$$

Alltså med sannolikhet 99% är

$$\frac{9\,S^2}{23.6} \leq \sigma^2 \leq \frac{9\,S^2}{1.73}.$$

Med skattningen $s^2 = 0.017422$ fås intervallet

$$0.0066471 = \frac{9\,s^2}{23.6} \leq \sigma^2 \leq \frac{9\,s^2}{1.73} = 0.090378 \qquad (99\%)$$

eller

$$0.08153 = \sqrt{0.0066471} \leq \sigma \leq \sqrt{0.090378} = 0.30063 \qquad (99\%).$$

13.20 Låt X_{ij} beskriva resultatet av den j:te mätningen i mätserie i. Modell:

$$X_{ij} = \text{fysikalisk konstant} + \text{mätfel} = m_i + \epsilon_{ij}$$

där ϵ_{ij} är oberoende $\mathrm{N}(0, \sigma)$, dvs X_{ij} är oberoende $\mathrm{N}(m_i, \sigma)$.

Alternativ 1: Observationer x_{ij} på X_{ij} i 8 mätserier med 2 observationer i varje ger upphov till $n = 8$ skattningar $s_1, \dots, s_8$ av mätmetodens standardavvikelse σ.

Mätserie, i	1	2	3	4	5	6	7	8
Antal obs., n_i	2	2	2	2	2	2	2	2
Medelvärde, $\overline{x}_i$	6.25	8.55	3.45	8.95	9.40	6.70	4.50	10.75
s_i	0.21213	0.21213	0.070711	0.35355	0.42426	0.28284	0.28284	0.35355

Som den sammanpoolade stickprovsvariansen fås

$$s^2 = \frac{(n_1 - 1)s_1^2 + \cdots + (n_8 - 1)s_8^2}{n_1 + \cdots + n_8 - 8} = \frac{0.045 + \cdots + 0.125}{8} = 0.085625,$$

dvs $s = 0.29262$ är en skattning av σ.

Vi vet att

$$\frac{8\,S^2}{\sigma^2} \quad \text{är} \quad \chi^2(8)\text{-fördelad}$$

där S^2 är stickprovsvariabeln motsvarande s^2. Alltså är

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{8\,S^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Med $\alpha = 0.95$ fås ur $\chi^2(8)$-tabeller

$$P\left(2.18 \leq \frac{8\,S^2}{\sigma^2} \leq 17.5\right) = 0.95.$$

Alltså med sannolikhet 95% är

$$\frac{8\,S^2}{17.5} \leq \sigma^2 \leq \frac{8\,S^2}{2.18}.$$

Med skattningen $s^2 = 0.085625$ fås intervallet

$$0.039066 = \frac{8\,s^2}{17.5} \leq \sigma^2 \leq \frac{8\,s^2}{2.18} = 0.31426 \qquad (95\%)$$

eller

$$0.19765 = \sqrt{0.039066} \leq \sigma \leq \sqrt{0.31426} = 0.56059 \qquad (95\%).$$

Alternativ 2: Differenserna

$$Y_i = X_{i1} - X_{i2}$$

är normalfördelade med väntevärde $E\,(Y_i) = E\,(X_{i1} - X_{i2}) = m_i - m_i = 0$ och varians $V\,(Y_i) = V\,(X_{i1} - X_{i2}) = \sigma^2 + \sigma^2 = 2\sigma^2$, dvs

$$Y_i \quad \text{är} \quad \mathrm{N}(0, \sqrt{2}\sigma).$$

Med observationer x_{ij} får vi

Mätserie, i	1	2	3	4	5	6	7	8
y_i	0.3	0.3	0.1	−0.5	−0.6	0.4	0.4	0.5

En skattning av $D\,(Y) = \sqrt{2}\sigma$ ges naturligtvis av $s_y = 0.42573$, dvs $s_y/\sqrt{2} = 0.30104$ är en skattning av σ, men bättre är att utnyttja att $E\,(Y) = 0$. En skattning av

$$2\sigma^2 = V\,(Y) = E\left((Y - E\,(Y))^2\right) = E\left(Y^2\right)$$

ges av

$$2\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2.$$

Notera att

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(Y_i^2\right) = \frac{1}{n}\sum_{i=1}^{n}(V\,(Y_i) + E\,(Y_i)^2) = 2\sigma^2$$

så $2\sigma^{*2}$ är en väntevärdesriktig skattning av $2\sigma^2$ och σ^{*2} är en väntevärdesriktig skattning av σ^2. Vi observerar

$$\sigma^{*2} = \frac{1}{2n}\sum_{i=1}^{n} y_i^2 = 0.085625$$

dvs $\sigma^* = 0.29262$ som skattning av σ. Nu är

$$\frac{n}{\sigma^2}\sigma^{*2} = \frac{1}{2\sigma^2}\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n}\left(\frac{Y_i - 0}{\sqrt{2}\sigma}\right)^2 = \sum_{i=1}^{n}\left(\mathrm{N}(0,1)\right)^2$$

en $\chi^2(n) = \chi^2(8)$-fördelad stokastisk variabel. Alltså är

$$P\left(\chi^2_{1-\alpha/2} \le \frac{8\,\sigma^{*2}}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Med $\alpha = 0.95$ fås ur $\chi^2(8)$-tabeller

$$P\left(2.18 \le \frac{8\,\sigma^{*2}}{\sigma^2} \le 17.5\right) = 0.95.$$

Alltså med sannolikhet 95% är

$$\frac{8\,\sigma^{*2}}{17.5} \le \sigma^2 \le \frac{8\,\sigma^{*2}}{2.18}.$$

Med skattningen $\sigma^{*2} = 0.085625$ fås intervallet

$$0.039066 = \frac{8\,\sigma^{*2}}{17.5} \le \sigma^2 \le \frac{8\,\sigma^{*2}}{2.18} = 0.31426 \qquad (95\%)$$

eller

$$0.19765 = \sqrt{0.039066} \le \sigma \le \sqrt{0.31426} = 0.56059 \qquad (95\%).$$

Alternativ 3: (Sämst.) Om man i alternativ 2 använder $s_y = 0.42573$ som en skattning av $\sqrt{2}\sigma$ kan man få ett konfidensintervall för σ genom att

$$\frac{(n-1)S_y^2}{2\sigma^2}$$

är en $\chi^2(n-1) = \chi^2(7)$-fördelad stokastisk variabel. Alltså är

$$P\left(\chi^2_{1-\alpha/2} \le \frac{7\,S_y^2}{2\sigma^2} \le \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Med $\alpha = 0.95$ fås ur $\chi^2(7)$-tabeller

$$P\left(1.69 \le \frac{7\,S_y^2}{2\sigma^2} \le 16.0\right) = 0.95.$$

Alltså med sannolikhet 95% är

$$\frac{7\,S_y^2}{2 \cdot 16.0} \le \sigma^2 \le \frac{7\,S_y^2}{2 \cdot 1.69}.$$

Med skattningen $s_y = 0.42573$ fås intervallet

$$0.039617 = \frac{7\,s_y^2}{2 \cdot 16.0} \le \sigma^2 \le \frac{7\,s_y^2}{2 \cdot 1.69} = 0.3754 \qquad (95\%).$$

eller

$$0.19904 = \sqrt{0.039617} \le \sigma \le \sqrt{0.3754} = 0.6127 \qquad (95\%).$$

13.21 Låt $X_1, \ldots, X_{n_x}$ beskriva den uppmätta överhöjningen för betongelement från fabrik A och $Y_1, \ldots, Y_{n_y}$ motsvarande för fabrik B. Modell: alla stokastiska variabler är oberoende och X_i är $\mathrm{N}(m_x, \sigma)$ och Y_i är $\mathrm{N}(m_y, \sigma)$. Då är

$$\frac{(\overline{X} - \overline{Y}) - (m_x - m_y)}{\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

$\mathrm{N}(0,1)$-fördelad och

$$\frac{(\overline{X} - \overline{Y}) - (m_x - m_y)}{S\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

$t(n_x + n_y - 2)$-fördelad, där

$$S = \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}}$$

är skattningsvariabeln för σ som utnyttjar både S_x och S_y. Alltså är med sannolikhet $1 - \alpha$

$$m_x - m_y \in \overline{X} - \overline{Y} \pm t_{\alpha/2} S \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}.$$

Med observationer

$$\overline{x} = 18.1,\ s_x = 5.0,\ n_x = 9 \qquad \overline{y} = 14.6,\ s_y = 7.1,\ n_y = 16$$

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

$$m_x - m_y \in 18.1 - 14.6 \pm 2.8073 \cdot 6.4476 \sqrt{\frac{1}{9} + \frac{1}{16}} = 3.5 \pm 7.54 \qquad (99\%)$$

eller intervallet $[-4.04,\ 11.0]$.

Enligt konfidensintervallet är 0 inte ett orimligt värde på $m_x - m_y$ varför vi inte kan utesluta $m_x = m_y$, dvs. det föreligger ingen påvisbar skillnad i förväntad överhöjning på nivå 1%.

13.22 Två oberoende stickprov. Låt $X_1, \ldots, X_n$ beskriva viktmätningarna för kolven utan vätska. Modell:

$$X_i = \underbrace{\text{kolvvikt}}_{=m_x} + \text{mätfel} = m_x + \epsilon_i$$

där ϵ_i är oberoende och $\mathrm{N}(0, \sigma)$, dvs X_i är $\mathrm{N}(m_x, \sigma)$. Låt $Y_1, \ldots, Y_r$ beskriva viktmätningarna för kolven med vätska. Här är modellen att $Y_1, \ldots, Y_r$ är oberoende $\mathrm{N}((\text{kolvvikt} + \text{vätska}), \sigma) = \mathrm{N}(m_y, \sigma)$.

Observationerna $x_1, \ldots, x_4$ och $y_1, \ldots, y_4$ ger skattningarna $\overline{x} = 15.04$ och $\overline{y} = 26.65$ av m_x respektive m_y. Vidare så är

$$s_x = 0.0150 \qquad s_y = 0.0096$$

två skattningar av samma σ. Båda dessa skattningar utnyttjas och den poolade skattningen av σ är

$$s = \sqrt{\frac{(n-1)s_x^2 + (r-1)s_y^2}{(n-1) + (r-1)}} = \sqrt{\frac{s_x^2 + s_y^2}{2}} = 0.012583.$$

Skattningen $\overline{y} - \overline{x}$ av $m_y - m_x$ beskrivs av

$$\overline{Y} - \overline{X} \text{ som är N} \left(m_y - m_x, \sigma \sqrt{\frac{1}{r} + \frac{1}{n}} \right).$$

Alltså,

$$\frac{(\overline{Y} - \overline{X}) - (m_y - m_x)}{\sigma \sqrt{\frac{1}{r} + \frac{1}{n}}} \text{ är N}(0,1) \qquad \text{och} \qquad \frac{(\overline{Y} - \overline{X}) - (m_y - m_x)}{S \sqrt{\frac{1}{r} + \frac{1}{n}}} \text{ är } t(n + r - 2)\text{-fördelad.}$$

Med sannolikhet $1 - \alpha$ så är

$$m_y - m_x \in (\overline{Y} - \overline{X}) \pm t_{\alpha/2} S \sqrt{\frac{1}{r} + \frac{1}{n}}.$$

Med konfidensgrad $1-\alpha = 0.95$ fås ur $t(n+r-2) = t(6)$-tabeller att $t_{\alpha/2} = t_{0.025} = 2.4469$ så intervallet blir

$$m_y - m_x \in (26.65 - 15.04) \pm 2.4469 \cdot 0.012583\sqrt{\frac{1}{4}+\frac{1}{4}} = 11.61 \pm 0.0218 \quad (95\%).$$

13.23 Låt $X_1, \ldots, X_n$ beskriva mätresultaten för proverna från det första partiet och $Y_1, \ldots, Y_r$ motsvarande för det andra partiet, med observerade värden $x_1, \ldots, x_n$ resp. $y_1, \ldots, y_r$. Modell: Alla stokastiska variabler är oberoende och

$$X_i \text{ är } \mathrm{N}(m_x, \sigma) \qquad Y_i \text{ är } \mathrm{N}(m_y, \sigma).$$

En skattning av m_x är $\overline{x}$ och en skattning av m_y är $\overline{y}$.

En skattning av $(m_x + m_y)/2$ ges av $(\overline{x} + \overline{y})/2$ som beskrivs av $(\overline{X} + \overline{Y})/2$ som har

$$\begin{aligned}
E\left(\frac{\overline{X}+\overline{Y}}{2}\right) &= E\left(\frac{1}{2n}\sum_{i=1}^{n} X_i + \frac{1}{2r}\sum_{j=1}^{r} Y_j\right) = \frac{1}{2n}\sum_{i=1}^{n} E\left(X_i\right) + \frac{1}{2r}\sum_{j=1}^{r} E\left(Y_j\right) \\
&= \frac{1}{2n}\sum_{i=1}^{n} m_x + \frac{1}{2r}\sum_{j=1}^{r} m_y = \frac{m_x + m_y}{2}.
\end{aligned}$$

och

$$\begin{aligned}
V\left(\frac{\overline{X}+\overline{Y}}{2}\right) &= V\left(\frac{1}{2n}\sum_{i=1}^{n} X_i + \frac{1}{2r}\sum_{j=1}^{r} Y_j\right) = \{\text{oberoende}\} \\
&= \frac{1}{(2n)^2}\sum_{i=1}^{n} V\left(X_i\right) + \frac{1}{(2r)^2}\sum_{j=1}^{r} V\left(Y_j\right) = \frac{1}{4n^2}\sum_{i=1}^{n}\sigma^2 + \frac{1}{4r^2}\sum_{j=1}^{r}\sigma^2 \\
&= \frac{\sigma^2}{4n} + \frac{\sigma^2}{4r}.
\end{aligned}$$

Alltså är

$$\frac{\overline{X}+\overline{Y}}{2} \quad \mathrm{N}\left(\frac{m_x+m_y}{2}, \frac{\sigma}{2}\sqrt{\frac{1}{n}+\frac{1}{r}}\right)$$

och

$$\frac{\frac{\overline{X}+\overline{Y}}{2} - \frac{m_x+m_y}{2}}{\frac{\sigma}{2}\sqrt{\frac{1}{n}+\frac{1}{r}}} \text{ är } \mathrm{N}(0,1)$$

och

$$\frac{\frac{\overline{X}+\overline{Y}}{2} - \frac{m_x+m_y}{2}}{\frac{S}{2}\sqrt{\frac{1}{n}+\frac{1}{r}}} \text{ är } t(n+r-2)$$

där

$$S^2 = \frac{(n-1)S_x^2 + (r-1)S_y^2}{n+r-2}.$$

Alltså med sannolikhet $1-\alpha$ är

$$-t_{\alpha/2}(n+r-2) \leq \frac{\frac{\overline{X}+\overline{Y}}{2} - \frac{m_x+m_y}{2}}{\frac{S}{2}\sqrt{\frac{1}{n}+\frac{1}{r}}} \leq t_{\alpha/2}(n+r-2)$$

eller, omformat, med sannolikhet $1-\alpha$ är

$$\frac{m_x+m_y}{2} \in \frac{\overline{X}+\overline{Y}}{2} \pm t_{\alpha/2}(n+r-2)\frac{S}{2}\sqrt{\frac{1}{n}+\frac{1}{r}}$$

Med data fås skattningen av $(m_x + m_y)/2$ till

$$\frac{\overline{x} + \overline{y}}{2} = \frac{24.2 + 28.5}{2} = 26.35$$

och skattningen av σ^2 fås till

$$s^2 = \frac{(5-1)1.6^2 + (10-1)2.3^2}{15-2} = 4.45.$$

Med konfidensgrad $1 - \alpha = 0.90$ fås $t_{\alpha/2}(n + r - 2) = t_{0.05}(13) = 1.77$ och

$$\frac{m_x + m_y}{2} \in \frac{24.2 + 28.5}{2} \pm 1.77 \frac{\sqrt{4.45}}{2} \sqrt{\frac{1}{5} + \frac{1}{10}} = 26.35 \pm 1.0231 \qquad (90\%)$$

eller $25.33 \leq \frac{m_x + m_y}{2} \leq 27.37$ (90%).

13.24 Låt $X_1, \ldots, X_n$ och $Y_1, \ldots, Y_n$ beskriva förslitningarna av däcktyp A resp. B på bil $1, 2, \ldots, n$. Modell: förslitningsskillnaderna

$$W_i = X_i - Y_i, \quad i = 1, \ldots, n$$

är oberoende och $\mathrm{N}(m, \sigma)$-fördelade stokastiska variabler. Observationerna

x_i :	1.0	0.9	0.7	1.5	0.5
y_i :	0.9	0.7	0.8	1.2	0.5
$w_i = x_i - y_i$:	0.1	0.2	−0.1	0.3	0

sammanfattas med

$$\overline{w} = \frac{1}{n} \sum_{i=1}^{n} w_i = 0.10 \qquad s_w^2 = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \overline{w})^2 = 0.025$$

så $s_w = \sqrt{0.025} = 0.1581$. Nu är $\overline{W}$ $\mathrm{N}(m, \sigma/\sqrt{n})$ och

$$\frac{\overline{W} - m}{S_w/\sqrt{n}} \text{ är } t(n-1).$$

Således, med sannolikhet $1 - \alpha$ är

$$-t_{\alpha/2} \leq \frac{\overline{W} - m}{S_w/\sqrt{n}} \leq t_{\alpha/2},$$

vilket omskrivet ger att med sannolikhet $1 - \alpha$ är

$$\overline{W} - t_{\alpha/2} \frac{S_w}{\sqrt{n}} \leq m \leq \overline{W} + t_{\alpha/2} \frac{S_w}{\sqrt{n}}.$$

Med våra $n = 5$ observationer 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

$$m \in \overline{w} \pm t_{\alpha/2} \frac{s_w}{\sqrt{n}} = 0.10 \pm 0.1963 = [-0.0963,\ 0.2963] \qquad (95\%).$$

13.25 Låt $x_1, \ldots, x_n$ vara bestämningar av blodtrycken för personer utan medicinering och $y_1, \ldots, y_r$ motsvarande för personer med medicinering. Modell: $x_1, \ldots, x_n$ och $y_1, \ldots, y_r$ är utfall av $X_1, \ldots, X_n$ resp $Y_1, \ldots, Y_r$, oberoende stokastiska variabler där

$$X_i \text{ är } \mathrm{N}(m_x, \sigma) \qquad Y_i \text{ är } \mathrm{N}(m_y, \sigma).$$

Med observationer

$$n = 50,\ \overline{x} = 148.2,\ s_x = 10.0 \qquad r = 25,\ \overline{y} = 151.7,\ s_y = 8.0$$

skattas σ med

$$s = \sqrt{\frac{(n-1)s_x^2 + (r-1)s_y^2}{(n-1)+(r-1)}} = \sqrt{88.164} = 9.3896.$$

Skillnaden $m_y - m_x$ skattas med $\overline{y} - \overline{x} = 3.5$ som beskrivs av

$$\overline{Y} - \overline{X} \text{ som är N}\left(m_y - m_x, \sigma\sqrt{\frac{1}{r} + \frac{1}{n}}\right).$$

Alltså är

$$\frac{(\overline{Y} - \overline{X}) - (m_y - m_x)}{\sigma\sqrt{\frac{1}{r} + \frac{1}{n}}} \quad \text{N}(0,1)$$

och

$$\frac{(\overline{Y} - \overline{X}) - (m_y - m_x)}{S\sqrt{\frac{1}{r} + \frac{1}{n}}} \quad \text{är } t(n + r - 2)\text{-fördelad.}$$

Ett konfidensintervall med konfidensgrad $1 - \alpha$ för $m_y - m_x$ ges av

$$m_y - m_x \in \overline{y} - \overline{x} \pm t_{\alpha/2} s\sqrt{\frac{1}{r} + \frac{1}{n}}.$$

Med $1 - \alpha = 0.95$ fås $t_{\alpha/2}(n + r - 2) = t_{0.025}(73) = 1.993$ och konfidensintervallet blir

$$m_y - m_x \in 151.7 - 148.2 \pm 1.993 \cdot 9{,}3896\sqrt{\frac{1}{25} + \frac{1}{50}} = 3.5 \pm 4.5838 = [-1.0838,\ 8.0838] \quad (95\%).$$

Enligt konfidensintervallet är $m_y - m_x = 0$ inte ett orimligt värde och vi kan inte utesluta att medicinen lämnar blodtrycket oförändrat.

Med stickprov i par. Låt $X_1, \ldots, X_n$ och $Y_1, \ldots, Y_n$ beskriva blodtrycken för patienter $1, \ldots, n$ före respektive efter medicinering. Modell:

$$Y_i = X_i + W_i, \quad i = 1, \ldots, n$$

där W_i är oberoende och N(m, σ)-fördelade stokastiska variabler. Det vill säga förändringarna $W_i = Y_i - X_i$ är oberoende och likafördelade. De $n = 25$ observationerna sammanfattas av

$$\overline{w} = 1.9 \quad s_w = 1.6.$$

Ett konfidensintervall för m med konfidensgrad $1 - \alpha$ ges av

$$m \in \overline{w} \pm t_{\alpha/2}(n-1)\frac{s_w}{\sqrt{n}}.$$

Med $1 - \alpha = 0.95$ är $t_{\alpha/2}(n - 1) = t_{0.025}(24) = 2.0639$ och konfidensintervallet blir

$$m \in 1.9 \pm 2.0639\frac{1.6}{\sqrt{25}} = 1.9 \pm 0.66045 = [1.24,\ 2.56] \quad (95\%).$$

Enligt konfidensintervallet är $m = 0$ inte ett rimligt värde så vi förkastar en hypotes om att medicinen inte förändrar blodtrycket på nivå 5%.

13.26 Låt $X_1, \ldots, X_n$ beskriva mätresultaten från laborant A. Eftersom proverna kan variera mellan olika dagar kan vi inte modellera mätresultaten som likafördelade stokastiska variabler. Låt $Y_1, \ldots, Y_n$ beskriva mätresultaten från laborant B. Modell: skillnaden i mätresultat:

$$W_i = Y_i - X_i \quad \text{är N}(m, \sigma),$$

$i = 1, \ldots, n$. Våra observationer är:

Dag nr:	1	2	3	4	5	6	7	8
x_i :	10	8	7	15	25	18	16	5
y_i :	11	14	6	19	24	21	22	8
$w_i = y_i - x_i$:	1	6	-1	4	-1	3	6	3

Data ger oss skattningarna

$$\overline{w} = 2.625 \qquad s = 2.7742$$

av m och σ.

Ett konfidensintervall för den systematiska skillnaden mellan laboranterna, m, ges av

$$m \in \overline{w} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}.$$

Med konfidensgrad $1 - \alpha = 0.95$ fås ur $t(n-1) = t(7)$-fördelningen att $t_{\alpha/2} = t_{0.025} = 2.3646$. Detta ger oss konfidensintervallet

$$m \in 2.625 \pm 2.3646 \cdot \frac{2.7742}{\sqrt{8}} = 2.6 \pm 2.3 \qquad (95\%)$$

13.27 Låt $X_1, \ldots, X_n$ vara de $n = 36$ stokastiska variabler som beskriver mätningar på tryckhållfastheten av betongbalkar med observerade utfall $x_1, \ldots, x_n$. Modell: $X_1, \ldots, X_n$ är oberoende och likafördelade där $E(X_i) = m$ och $V(X_i) = \sigma^2$, $i = 1, \ldots, n$.

Observationerna klassindelas och resultatet redovisas i tabellen nedan:

Klassmitt	1200	1400	1600	1800	2000	2200	2400	2600	2800	[kg]
Antal obs.	1	3	2	7	9	6	5	1	2	

En skattning av m är $\overline{x}$ där

$$\overline{x} \approx \frac{1}{36} (1200 \cdot 1 + 1400 \cdot 3 + \cdots + 2800 \cdot 2) = 2016.7$$

En skattning av σ^2 är

$$s^2 \approx \frac{1}{35} \left((1200 - \overline{x})^2 \cdot 1 + (1400 - \overline{x})^2 \cdot 3 + \cdots + (2800 - \overline{x})^2 \cdot 2 \right) = 1.4257 \cdot 10^5$$

så $s = 377.59$ är en skattning av σ. Enligt Centrala gränsvärdessatsen är

$$\sum_{i=1}^{n} X_i$$

approximativt normalfördelad vilket gör att

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{är approximativt N} \left(m, \frac{\sigma}{\sqrt{n}} \right).$$

Ett konfidensintervall av konfidensgrad approximativt $1 - \alpha = 0.99$ för m ges av

$$m \in \overline{x} \pm \lambda_{\alpha/2} \frac{s}{\sqrt{n}} = 2016.7 \pm 2.5758 \frac{377.59}{\sqrt{36}} = 2016.7 \pm 162.1 \qquad (\approx 99\%)$$

Om konfidensintervallet skall ha bredden 100, dvs vara $\overline{x} \pm 50$ så skall

$$50 = \lambda_{\alpha/2} \frac{s}{\sqrt{n}}$$

eller $n = (\lambda_{\alpha/2}s/50)^2 = 378.38$, dvs minst 379 observationer krävs.

13.28 Låt $X_1, \ldots, X_n$ beskriva de uppmätta föroreningshalterna uppströms, och $Y_1, \ldots, Y_r$ motsvarande nedströms.

Modell: Alla stokastiska variabler är oberoende och $X_1, \ldots, X_n$ är likafördelade med $E(X_i) = m_x$ och $V(X_i) = \sigma_x^2$ och $Y_1, \ldots, Y_r$ är likafördelade med $E(Y_i) = m_y$ och $V(Y_i) = \sigma_y^2$. Med utfall $x_1, \ldots, x_n$ av $X_1, \ldots, X_n$ och $y_1, \ldots, y_r$ av $Y_1, \ldots, Y_r$ observerades följande skattningar:

$$\begin{array}{rlll} \text{Uppströms:} & n = 30, & \overline{x} = 13.2, & s_x = 2.80 \\ \text{Nedströms:} & r = 40, & \overline{y} = 86.1, & s_y = 38.7 \end{array}$$

Enligt Centrala gränsvärdessatsen är

$$\sum_{i=1}^{n} X_i \quad \text{och} \quad \sum_{i=1}^{r} Y_i$$

approximativt normalfördelade vilket gör att

$$\overline{Y} - \overline{X} = \frac{1}{r}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{r} X_i \quad \text{är approximativt N}\left(m_y - m_x, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{r}}\right).$$

Ett konfidensintervall av konfidensgrad approximativt $1 - \alpha = 0.95$ för $m_y - m_x$ ges av

$$m_y - m_x \in \overline{y} - \overline{x} \pm \lambda_{\alpha/2}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{r}} = 72.9 \pm 12.035 \qquad (\approx 95\%)$$

eller $60.865 \leq m_y - m_x \leq 84.935$ ($\approx 95\%$).

13.29 Låt $X_1, \ldots, X_n$ beskriva antalet bilar som passerar en viss gata under n olika dagar före en trafiksanering, och $Y_1, \ldots, Y_r$ motsvarande efter en trafiksanering.

Modell: Alla stokastiska variabler är oberoende och $X_1, \ldots, X_n$ är likafördelade med $E(X_i) = m_x$ och $V(X_i) = \sigma_x^2$ och $Y_1, \ldots, Y_r$ är likafördelade med $E(Y_i) = m_y$ och $V(Y_i) = \sigma_y^2$. Med utfall $x_1, \ldots, x_n$ av $X_1, \ldots, X_n$ och $y_1, \ldots, y_r$ av $Y_1, \ldots, Y_r$ observerades följande skattningar:

$$\begin{array}{rlll} \text{Före:} & n = 30, & \overline{x} = 6342, & s_x = 108 \\ \text{Efter:} & r = 20, & \overline{y} = 956, & s_y = 24 \end{array}$$

Enligt Centrala gränsvärdessatsen är

$$\sum_{i=1}^{n} X_i \quad \text{och} \quad \sum_{i=1}^{r} Y_i$$

approximativt normalfördelade vilket gör att

$$\overline{X} - \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{r}\sum_{i=1}^{r} Y_i \quad \text{är approximativt N}\left(m_x - m_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{r}}\right).$$

Ett konfidensintervall av konfidensgrad approximativt $1 - \alpha = 0.95$ för $m_x - m_y$ ges av

$$m_x - m_y \in \overline{x} - \overline{y} \pm \lambda_{\alpha/2}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{r}} = 5386 \pm 40.053 \qquad (\approx 95\%)$$

eller $5345.9 \leq m_x - m_y \leq 5426.1$ ($\approx 95\%$).

13.30 Låt X beskriva antalet felaktiga enheter av $n = 500$ undersökta. Modell: en vald enhet är felaktig med sannolikhet p oberoende av andra enheter. Då är X binomialfördelad och en skattningsvariabel av p är den relativa frekvensen $p^* = X/n$ med observerat utfall $x/n = 87/500 = 0.174$. Nu är

$$E(p^*) = E\left(\frac{X}{n}\right) = \frac{1}{n}\underbrace{E(X)}_{=np} = \frac{1}{n}np = p$$

och

$$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}\underbrace{V(X)}_{=np(1-p)} = \frac{p(1-p)}{n}.$$

En skattning av $V(X) = np(1-p)$ ges av $np^*(1-p^*)$ med observerat värde $500 \cdot 0.174(1-0.1674) = 71.862$ vilket är större än 10 med råge. Därför borde X:s binomialfördelning kunna approximeras med normalfördelningen. Då är även p^* approximativt normalfördelad, dvs p^* är approximativt $\mathrm{N}(p, \sqrt{p(1-p)/n})$. Med sannolikhet approximativt $1-\alpha$ är således

$$-\lambda_{\alpha/2} \leq \frac{p^* - p}{\sqrt{p(1-p)/n}} \leq \lambda_{\alpha/2}$$

vilket omformat ger konfidensintervallet

$$p \in p^* \pm \lambda_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

som approximeras med

$$p \in p^* \pm \lambda_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n}}.$$

Med konfidensgrad $1-\alpha = 0.95$ fås $\lambda_{0.025} = 1.9600$ och

$$p \in 0.174 \pm 0.03323 = [0.14077,\ 0.20723] \qquad (\approx 95\%).$$

För det minsta av de rimliga värdena på p, 0.14077, fås variansskattningen av X till $500 \cdot 0.14077(1-0.14077) = 60.477$ vilket även det är större än 10 och normalapproximationen är giltig.

13.31 Låt X beskriva det antal gånger maskinen stod stilla vid $n = 250$ observationer. Modell: maskinen står still vid ett observationstillfälle med sannolikhet p oberoende av andra tillfällen. Då är X binomialfördelad och en skattningsvariabel av p är den relativa frekvensen $p^* = X/n$ med observerat utfall $x/n = 42/250 = 0.168$. Nu är

$$E(p^*) = E\left(\frac{X}{n}\right) = \frac{1}{n}\underbrace{E(X)}_{=np} = \frac{1}{n}np = p$$

och

$$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}\underbrace{V(X)}_{=np(1-p)} = \frac{p(1-p)}{n}.$$

En skattning av $V(X) = np(1-p)$ ges av $np^*(1-p^*)$ med observerat värde $200 \cdot 0.168(1-0.168) = 34.94$ vilket är större än 10 med råge. Därför borde X:s binomialfördelning kunna approximeras med normalfördelningen. Då är även p^* approximativt normalfördelad, dvs p^* är approximativt $\mathrm{N}(p, \sqrt{p(1-p)/n})$. Med sannolikhet approximativt $1-\alpha$ är således

$$-\lambda_{\alpha/2} \leq \frac{p^* - p}{\sqrt{p(1-p)/n}} \leq \lambda_{\alpha/2}$$

vilket omformat ger konfidensintervallet

$$p \in p^* \pm \lambda_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

som approximeras med

$$p \in p^* \pm \lambda_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n}}.$$

Med konfidensgrad $1-\alpha = 0.95$ fås $\lambda_{0.025} = 1.9600$ och

$$p \in 0.168 \pm 0.0463 = [0.1217,\ 0.2143] \qquad (\approx 95\%).$$

För det minsta av de rimliga värdena på p, 0.1217, fås variansskattningen av X till $250 \cdot 0.1444(1-0.1444) = 26.71$ vilket även det är större än 10 och normalapproximationen är giltig.

13.32 Låt X vara $\text{Bin}(n, p)$ där som skattningsvariabel av p används den relativa frekvensen $p^* = X/n$. Antag att $V(X)$ är så stor att normalapproximation av binomialfördelningen är tillämplig. Då är p^* är approximativt $\text{N}(p, \sqrt{p(1-p)/n})$. Med sannolikhet approximativt $1-\alpha$ är således

$$-\lambda_{\alpha/2} \leq \frac{p^* - p}{\sqrt{p(1-p)/n}} \leq \lambda_{\alpha/2}$$

vilket omformat ger konfidensintervallet

$$p \in p^* \pm \lambda_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = p^* \pm \Delta(p)$$

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

$$\frac{d}{dp}\left[p(1-p)\right] = 1 - 2p = 0$$

ger $p = 1/2$ och kontroll av andraderivatan 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.

Med $\Delta = 0.05$ och $1 - \alpha = 0.95$ fås $\lambda_{\alpha/2} = 1.9600$ och

$$n = \frac{1.96^2}{4 \cdot 0.05^2} = 384.16$$

dvs minst 385 observationer.

Med kunskap om att $p \leq 0.10$ kan man utnyttja att $p(1-p)$ är växande på intervallet $[0, \frac{1}{2})$ och

$$\Delta(p) \leq \lambda_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\Bigg|_{p=0.10} = \lambda_{\alpha/2}\sqrt{\frac{0.09}{n}} = \Delta.$$

Omformat fås att

$$n = \frac{\lambda_{\alpha/2}^2 0.09}{\Delta^2} = \frac{1.96^2 \cdot 0.09}{0.05^2} = 138.3$$

dvs 139 observationer ger ett konfidensintervall vars halva bredd $\Delta(p)$ är högst $\Delta = 0.05$ då $p \leq 0.10$.

13.33 Låt X_A beskriva antalet kasserade enheter av $n_A = 100$ tillverkade av maskin A och X_B motsvarande för $n_B = 200$ tillverkade av maskin B. Modell: en vald enhet från maskin A eller B är felaktig med sannolikhet p_A resp. p_B oberoende av andra tillverkade enheter. Då är X_A och X_B oberoende och binomialfördelade och skattningsvariablerna av p_A resp. p_B är $p_A^* = X_A/n_A$ och $p_B^* = X_B/n_B$ med observerade utfall $x_A/n_A = 36/100 = 0.36$ resp. $x_B/n_B = 56/200 = 0.28$. Nu är

$$E\left(p_A^* - p_B^*\right) = E\left(\frac{X_A}{n_A} - \frac{X_B}{n_B}\right) = \frac{1}{n_A}\underbrace{E\left(X_A\right)}_{=n_A p_A} - \frac{1}{n_B}\underbrace{E\left(X_B\right)}_{=n_B p_B} = p_A - p_B$$

och

$$V\left(p_A^* - p_B^*\right) = V\left(\frac{X_A}{n_A} - \frac{X_B}{n_B}\right) = \frac{1}{n_A^2}V\left(X_A\right) + \frac{(-1)^2}{n_B^2}V\left(X_B\right) = \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}$$

En skattning av $V\left(X_A\right) = n_A p_A(1-p_A)$ ges av $n_A p_A^*(1-p_A^*)$ med observerat värde $100 \cdot 0.36(1-0.36) = 23.04$ vilket är större än 10 och fördelningen för X_A kan approximeras med normalfördelningen och då är även p_A^* approximativt normalfördelad. På samma sätt skattas $V\left(Y\right)$ till 40.32 så fördelningen för p_B^* är approximativt normal. Alltså

$$p_A^* - p_B^* \text{ är approximativt N}\left(p_A - p_B, \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}\right)$$

Ett konfidensintervall med konfidensgrad approximativt $1-\alpha$ för $p_A - p_B$ ges av

$$p_A - p_B \in p_A^* - p_B^* \pm \lambda_{\alpha/2}\sqrt{\frac{p_A^*(1-p_A^*)}{n_A} + \frac{p_B^*(1-p_B^*)}{n_B}}.$$

Med konfidensgrad $1-\alpha = 0.95$ fås $\lambda_{0.025} = 1.9600$ och

$$p_A - p_B \in 0.36 - 0.28 \pm 1.9600\sqrt{\frac{0.36(1-0.36)}{100} + \frac{0.28(1-0.28)}{200}} = 0.08 \pm 0.1128 \qquad (\approx 95\%).$$

Enligt detta konfidensintervall är 0 inte ett orimligt värde på skillnaden $p_A - p_B$ och man kan inte utesluta att $p_A = p_B$, dvs att maskinerna har samma kassationssannolikhet.

Med utfallet $x_A = 52$ fås istället $x_A/n_A = 0.52$ och en skattning av $V\left(X\right)$ till 24.96. Normalapproximation är fortfarande tillåten och det approximativa konfidensintervallet blir

$$\begin{aligned} p_A - p_B &\in 0.52 - 0.28 \pm 1.9600\sqrt{\frac{0.52(1-0.52)}{100} + \frac{0.28(1-0.28)}{200}} \qquad (\approx 95\%). \\ &= 0.24 \pm 0.1160 = [0.124,\ 0.356]. \end{aligned}$$

Enligt detta konfidensintervall är 0 ett orimligt värde på skillnaden $p_A - p_B$ och man kan utesluta att $p_A = p_B$ (på risknivå approximativt 5%).

13.34 Låt X beskriva antalet bränder vid n_1 olyckor av bilmärke A och Y motsvarande för n_2 olyckor av bilmärke B. Modell:

$$X \text{ är Bin}(n_1, p_1) \qquad Y \text{ är Bin}(n_2, p_2).$$

där X och Y är oberoende. Variansskattningar för X och Y avgör om normalapproximation av binomialfördelningarna är tillämplig. Då är $p_1^* = X/n_1$ och $p_2^* = Y/n_2$ också approximativt normalfördelade.

Alternativ 1: Ett konfidensintervall för $E\left(p_1^* - p_2^*\right) = p_1 - p_2$ med konfidensgrad approximativt $1-\alpha$ ges av

$$p_1 - p_2 \in p_1^* - p_2^* \pm \lambda_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}.$$

Om 0 ingår i intervallet är 0 ett rimligt värde på skillnaden $p_1 - p_2$ och man kan inte utesluta att $p_1 = p_2$, dvs att det inte föreligger någon skillnad i brandsannolikheter för bilmärke A och B på signifikansnivå (risknivå) α.

Alternativ 2: Vi vill testa

$$H_0 : p_1 = p_2 \text{ (alt. } p_1 - p_2 = 0) \quad \text{mot} \quad H_1 : p_1 \neq p_2 \text{ (alt. } p_1 - p_2 \neq 0)$$

på signifikansnivå (risknivå) α. Då H_0 är sann så är

$$p_1^* - p_2^* \text{ approximativt N}\left(0, \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}\right)$$

där $p = p_1 = p_2$ skattas med $p^* = (X+Y)/(n_1+n_2)$. Vi förkastar H_0 om

$$\left|\frac{p_1^* - p_2^*}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right| \geq \lambda_{\alpha/2}.$$

Med observationer

$$n_1 = 37653,\ p_1^* = 0.29 \qquad n_2 = 13245,\ p_2^* = 0.13.$$

Alternativ 1: Vi får konfidensintervallet med konfidensgrad approximativt $1 - \alpha = 0.95$ ges av

$$p_1 - p_2 \in 0.16 \pm 0.0073356 = [0.1527,\ 0.1673] \qquad (\approx 95\%)$$

Eftersom 0 inte ingår i intervallet kan vi utesluta att $p_1 = p_2$, dvs det föreligger en skillnad i brandsannolikheter för bilmärke A och B på signifikansnivå (risknivå) $\alpha = 5\%$.

Alternativ 2: Vi skattar det gemensamma p med

$$p^* = \frac{n_1 p_1^* + n_2 p_2^*}{n_1 + n_2} = 0.24836$$

och observerar utfallet

$$\left|\frac{p_1^* - p_2^*}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right| = 36.656 > 1.9600 = \lambda_{\alpha/2}$$

så vi förkastar H_0, hypotesen om samma brandsannolikhet för bilmärke A och B. Den lägsta signifikansnivån vi skulle förkasta H_0 för är

$$\text{p-värde} = P\left(\left|\frac{p_1^* - p_2^*}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right| > 36.656\right) \approx 2(1 - \Phi(36.656)) \approx 3.6 \cdot 10^{-294}.$$

13.35 Låt $X_1, \ldots, X_n$ beskriva antalet samtal till växeln under det aktuella tidsintervallet för olika dagar med observerade värden $x_1, \ldots, x_n$. Modell: $X_1, \ldots, X_n$ är oberoende och X_i är Poissonfördelad med parameter (väntevärde) m. Väntevärdet m skattas med

$$\overline{x} = 100.88 \text{ samtal.}$$

Skattningen beskrivs av

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

där

$$Y = X_1 + \cdots + X_n$$

är Poissonfördelad med väntevärde

$$E(Y) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = nm$$

som skattas med $\sum x_i = 807$ vilket är större än 15 med råge så Poissonfördelningen för Y kan approximeras med normalfördelning, men då är även $\overline{X}$ approximativt normalfördelad. Alltså,

$$\overline{X} \text{ är approximativt N}\left(m, \sqrt{\frac{m}{n}}\right)$$

Ett konfidensintervall för m med konfidensgrad approximativt $1-\alpha = 0.95$ ges av

$$m \in \overline{x} \pm \lambda_{\alpha/2}\sqrt{\frac{\overline{x}}{n}} = 100.88 \pm 6.9599 = [93.9,\ 107.8] \qquad (\approx 95\%).$$

13.36 Låt $X_1, \ldots, X_n$ beskriva antalet registrerade partiklar under tidsintervall av längd $t_1, \ldots, t_n$ med observerade värden $x_1, \ldots, x_n$. Modell: $X_1, \ldots, X_n$ är oberoende och X_i är Poissonfördelad med parameter (väntevärde) λt_i. Intensiteten λ skattas med

$$\lambda^* = \frac{x_1 + \cdots + x_n}{t_1 + \cdots + t_n} = 14.571 \text{ partiklar/minut.}$$

Skattningen beskrivs av

$$\Lambda^* = \frac{X_1 + \cdots + X_n}{t_1 + \cdots + t_n}$$

där

$$Y = X_1 + \cdots + X_n$$

är Poissonfördelad med väntevärde

$$E(Y) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = \lambda \sum_{i=1}^{n} t_i$$

som skattas med $\lambda^* \sum_{i=1}^{n} t_i = 1020$ vilket är större än 15 med råge så Poissonfördelningen för Y kan approximeras med normalfördelning, men då är även Λ^* approximativt normalfördelad. Alltså,

$$\Lambda^* \text{ är approximativt N}\left(\lambda, \sqrt{\frac{\lambda}{\sum_{i=1}^{n} t_i}}\right)$$

och ett konfidensintervall för λ med konfidensgrad approximativt $1-\alpha = 0.95$ ges av

$$\lambda \in \lambda^* \pm \lambda_{\alpha/2}\sqrt{\frac{\lambda^*}{\sum_{i=1}^{n} t_i}} = 14.571 \pm 0.894 \qquad (\approx 95\%)$$

13.37 Om X är Poissonfördelad med parameter λt, X är Po(λt), så är

$$E(X) = \lambda t = V(X).$$

Med $\lambda^* = X/t$ fås

$$E(\lambda^*) = E\left(\frac{1}{t}X\right) = \frac{1}{t}E(X) = \frac{1}{t}\lambda t = \lambda$$

och

$$V(\lambda^*) = V\left(\frac{1}{t}X\right) = \frac{1}{t^2}E(X) = \frac{1}{t^2}\lambda t = \frac{\lambda}{t}.$$

Notera att x/t är en väntevärdesriktig skattning av λ. Om $E(X) \geq 15$ så kan Poissonfördelningen approximeras med normalfördelningen, dvs X är approximativt N($\lambda t, \sqrt{\lambda t}$), dvs λ^* är approximativ N($\lambda, \sqrt{\lambda/t}$). Med sannolikhet approximativt $1-\alpha$ så är

$$-\lambda_{\alpha/2} \leq \frac{\lambda^* - \lambda}{\sqrt{\lambda/t}} \leq \lambda_{\alpha/2}$$

vilket omformas till

$$\lambda \in \lambda^* \pm \lambda_{\alpha/2}\sqrt{\frac{\lambda}{t}}$$

som approximeras med

$$\lambda \in \lambda^* \pm \lambda_{\alpha/2}\sqrt{\frac{\lambda^*}{t}}.$$

Med konfidensgrad $1-\alpha = 0.99$ fås $\lambda_{\alpha/2} = \lambda_{0.005} = 2.5758$. Vi skall bestämma t så att konfidensintervallets halva bredd

$$\lambda_{\alpha/2}\sqrt{\lambda/t} \leq 25,$$

dvs t skall uppfylla

$$t \geq \left(\frac{\lambda_{\alpha/2}}{25}\right)^2 \lambda$$

Nu är $\lambda \leq 1000$ [tim^{-1}] så med $t \geq (2.5758/25)^2 \cdot 1000 = 10.6158$ [timmar] är kraven uppfyllda.

Med $t = 10$, $x = 7835$ fås ett konfidensintervall med konfidensgrad approximativt $1 - \alpha = 0.99$

$$\lambda \in \frac{x}{t} \pm \lambda_{\alpha/2}\sqrt{\frac{x/t}{t}} = 783.5 \pm 22.8 = [760.70,\ 806.30] \qquad (\approx 99\%).$$

13.38 Låt $X_1, \ldots, X_n$ beskriva tiderna mellan fel hos den komplicerade tekniska utrustningen. Modell: $X_1, \ldots, X_n$ är oberoende och exponentialfördelade med parameter m,

$$E(X) = m \qquad V(X) = m^2,$$

och observerade värden $x_1, \ldots, x_n$. En skattning av $E(X) = m$ är $\overline{x}$ som beskrivs av $\overline{X}$ som enligt Centrala gränsvärdessatsen är approximativt normalfördelad om n är stort. Alltså är

$$\overline{X} \text{ approximativt N}\left(m, \frac{m}{\sqrt{n}}\right)$$

eller

$$\frac{\overline{X} - m}{m/\sqrt{n}} \text{ är approximativt N}(0,1)$$

så med sannolikhet $\approx 1 - \alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{\overline{X} - m}{m/\sqrt{n}} \leq \lambda_{\alpha/2}.$$

Olikheten

$$\frac{\overline{X} - m}{m/\sqrt{n}} \leq \lambda_{\alpha/2}$$

ger att

$$m \geq \frac{1}{1 + \lambda_{\alpha/2}/\sqrt{n}}\overline{X}.$$

På samma sätt fås att olikheten

$$-\lambda_{\alpha/2} \leq \frac{\overline{X} - m}{m/\sqrt{n}}$$

kan skrivas

$$m \leq \frac{1}{1 - \lambda_{\alpha/2}/\sqrt{n}}\overline{X}.$$

Alltså, med sannolikhet $\approx 1 - \alpha$ är

$$m \in \left[\underbrace{\frac{1}{1 + \lambda_{\alpha/2}/\sqrt{n}}}_{=k_1}\overline{X}, \quad \underbrace{\frac{1}{1 - \lambda_{\alpha/2}/\sqrt{n}}}_{=k_2}\overline{X}\right] = \left[k_1\overline{X}, k_2, \overline{X}\right].$$

Med $n = 25$ komponenter erhölls $\overline{x} = 1230$. Med konfidensgrad $1 - \alpha = 0.95$ fås $\lambda_{\alpha/2} = 1.9600$ och det observerade intervallet

$$m \in [0.71839\overline{x},\ 1.6447\overline{x}] = [883.63,\ 2023.0] \qquad (\approx 95\%).$$

För att få $k_1 = \frac{1}{1+\lambda_{\alpha/2}/\sqrt{n}} = 0.9$ krävs

$$n = \left(\frac{\lambda_{\alpha/2}}{\frac{1}{k_1} - 1}\right)^2 = \left(\frac{1.9600}{\frac{1}{0.9} - 1}\right)^2 = 311.17,$$

dvs. minst 312 komponenter.

Man kan sätta upp exakta konfidensintervall för m i exponentialfördelningen om man vet att

$$\frac{2}{m}\sum_{i=1}^{n} X_i$$

är en $\chi^2(2n)$-fördelad stokastisk variabel. Då kan man bestämma kvantilerna så att med sannolikhet $1-\alpha$ är

$$\chi^2_{1-\alpha/2} \leq \frac{2}{m}\sum_{i=1}^{n} X_i \leq \chi^2_{\alpha/2}$$

eller

$$\frac{2\sum_{i=1}^{n} X_i}{\chi^2_{\alpha/2}} \leq m \leq \frac{2\sum_{i=1}^{n} X_i}{\chi^2_{1-\alpha/2}}.$$

Med $n = 25$ fås ur $\chi^2(2n) = \chi^2(50)$-tabeller att $\chi^2_{0.975} = 32.357$ och $\chi^2_{0.025} = 71.42$ så

$$861.1 \leq m \leq 1900.6 \qquad (95\%).$$

13.39 Låt X beskriva antalet komponenter bland n som fungerar vid tiden T. Modell: X är $\mathrm{Bin}(n, p)$ med $p = \mathrm{e}^{-T/m}$. Som skattningsvariabel av p används den relativa frekvensen $p^* = X/n$. Antag att $V(X)$ är så stor att normalapproximation av binomialfördelningen är tillämplig. Då är p^* approximativt $\mathrm{N}(p, \sqrt{p(1-p)/n})$. Ett konfidensintervall för p med konfidensgrad approximativt $1 - \alpha$ ges av

$$p^* - \lambda_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n}} \leq p \leq p^* + \lambda_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n}}.$$

Med $x = 90, n = 100, T = 1000$ fås $p^* = 90/100 = 0.90$. Med konfidensgrad $1-\alpha = 0.95$ är $\lambda_{0.025} = 1.9600$ och intervallet blir

$$0.8412 \leq p = \mathrm{e}^{-T/m} \leq 0.9588 \qquad (\approx 95\%)$$

vilket ger

$$5782.8 = -T/\ln(0.8412) \leq m \leq -T/\ln(0.9588) = 23768 \qquad (\approx 95\%)$$

14.1 Låt X beskriva antalet par av $n = 18$ där det behandlade röret rostat mest. Modell: X är $\mathrm{Bin}(n, p)$. Vi vill testa

$$H_0 : p = \frac{1}{2} \qquad \text{mot} \qquad H_1 : p < \frac{1}{2}$$

genom att förkasta H_0 då $X \leq 5$. Signifikansnivån (risknivån) är

$$\alpha \quad = \quad P\,(\text{förkasta korrekt } H_0) = P\,(X \leq 5) = \sum_{k=0}^{5} \binom{n}{k} p^k (1-p)^{n-k}\Big|_{p=1/2} = 0.0481$$

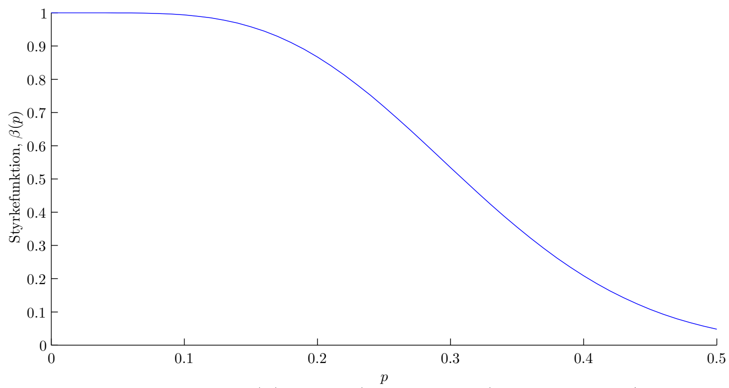
Slutsatsen om man ser utfallet…

- ... $x = 3 < 5$ 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

$$\beta(p) = P\left(\text{förkasta } H_0\right) = P\left(X \leq 5\right) = \sum_{k=0}^{5} \binom{n}{k} p^k (1-p)^{n-k}$$

för $p \leq 1/2$. Då är $\beta(0.25) = 0.7175$.



Styrkefunktionen $\beta(p) = P\left(\text{förkasta } H_0\right)$ för $p \leq 1/2$. Ju längre från $H_0 : p = 1/2$ desto större är sannolikheten att nollhypotesen förkastas.

14.2 I ett parti om N enheter väljs n på måfå utan återläggning. Partier innehåller pN defekta enheter och man godtar partiet om högst c av de utvalda är defekta. Låt X beteckna antalet defekta bland de utvalda. Modell: X är hypergeometriskt fördelat:

$$P\left(X = k\right) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$$

Om $n \ll N$ kan vi använda oss av binomialapproximation

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

Vi vill välja n och c så att

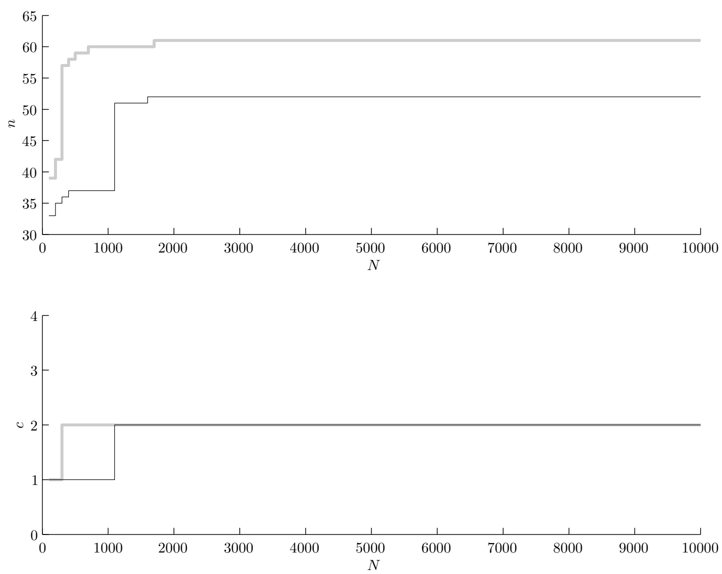
$$P\left(\text{Ej acceptera partiet}\right) = P\left(X > c\right) \leq 0.05$$

för $0 \leq p \leq 0.01$, och

$$P\left(\text{Ej acceptera partiet}\right) = P\left(X > c\right) \geq 0.90$$

för $0.05 \leq p \leq 1$. Binomialfördelningen ger att valet $n = 52$ och $c = 2$ medför

$$P\left(X > c\right)\big|_{p=0.01} = 0.01535 \quad \text{och} \quad P\left(X > c\right)\big|_{p=0.05} = 0.90337.$$



Visar valet av n (överst) och c (underst) som för olika populationsstorlekar N som uppfyller $P(X > c) \leq 0.05$ då $0 \leq p \leq 0.01$ och $P(X > c) \geq 0.90$ (svart) eller $P(X > c) \geq 0.95$ (grå), för $p \geq 0.10$.

14.3 Låt $X_1, \ldots, X_n$, $n = 30$, vara de stokastiska variabler som beskriver de observerade utfallen $x_1, \ldots, x_n$. Modell: X_i, $i = 1, \ldots, n$, är oberoende och $\mathrm{N}(m, \sigma)$ fördelade. Vi vill testa

$$H_0 : \sigma = 4 = \sigma_0 \quad \text{mot} \quad H_1 : \sigma < \sigma_0$$

på nivå $\alpha = 0.01$. Vi förkastar H_0 till förmån för H_1 för små värden på stickprovsvariansen S^2, alternativt för små värden på

$$\frac{(n-1)S^2}{\sigma_0^2}$$

som om H_0 är sann är en $\chi^2(n-1)$-fördelad stokastisk variabel. Så om H_0 är sann så är med sannolikhet α

$$\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha},$$

eller, omformat,

$$S^2 \leq \chi^2_{1-\alpha} \frac{\sigma_0^2}{n-1}.$$

Med signifikansnivå $\alpha = 0.01$ fås ur $\chi^2(n-1) = \chi^2(29)$-tabeller att $\chi^2_{1-\alpha} = \chi^2_{0.99} = 14.26$ så vi förkastar H_0 om vi observerar ett utfall

$$S^2 \leq \chi^2_{1-\alpha} \frac{\sigma_0^2}{n-1} = 14.26 \frac{4^2}{30-1} = 7.8656$$

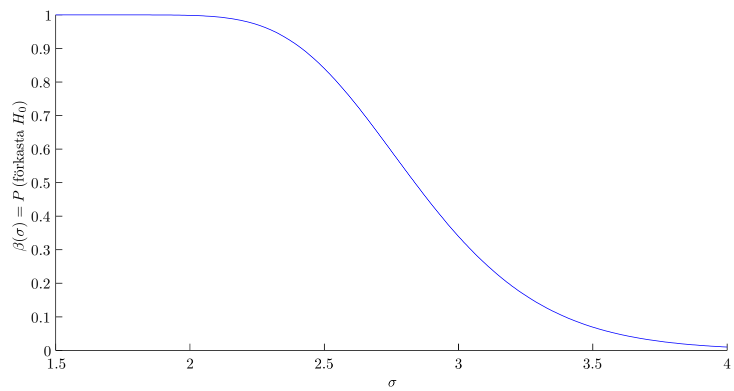
alternativt förkasta H_0 om $S \leq \sqrt{7.8656} = 2.8046$.

Med utfallet $s = 3.1 > 2.805$ så tillhör s inte förkastelseområdet och H_0 förkastas ej på nivå 1%. Om $\sigma = 2$ så är

$$\begin{aligned}
\beta(\sigma) &= P(\text{förkasta } H_0) = P\left(S^2 \leq \chi^2_{1-\alpha} \frac{\sigma_0^2}{n-1}\right) = P\left(\frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha} \frac{\sigma_0^2}{\sigma^2}\right) \\
&= P\left(\frac{(n-1)S^2}{\sigma^2} \leq 14.26 \frac{4^2}{\sigma^2}\right)
\end{aligned}$$

så, med hjälp av fördelningsfunktionen för en $\chi^2(29)$-fördelning,

$$\beta(2) = P\left(\frac{(n-1)S^2}{\sigma^2} \leq 57.0258\right) = 0.9986$$



Styrkefunktionen $\beta(\sigma)$

14.4 Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(m, \sigma)$ beskrivandes längden av plankorna. 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 13.15)

$$0.50623 = \sqrt{\frac{(n-1)s^2}{27.5}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{6.26}} = 1.0606 \qquad (95\%).$$

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

$$\frac{(n-1)S^2}{0.4^2} \quad \text{en} \quad \chi^2(n-1)\text{-fördelad stokastisk variabel.}$$

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 ett utfall $(n-1)s^2/0.4^2$ 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 $\alpha = 0.05$.

14.5 Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(m, \sigma)$ beskrivandes diametrarna på huvudena. Vi vill testa

$$H_0 : \sigma \leq 0.02 \quad \text{mot} \quad H_1 : \sigma > 0.02$$

på signifikansnivå $\alpha = 0.05$.

Vi förkastar H_0 då S är mycket större än 0.2, alternativt då

$$\frac{(n-1)S^2}{0.02^2}$$

är väldigt stor. Då $\sigma = 0.02$ (och H_0 är sann) så är

$$\frac{(n-1)S^2}{0.02^2} \quad \text{en} \quad \chi^2(n-1)\text{-fördelad stokastisk variabel.}$$

Alltså, då $\sigma = 0.02$ är

$$P\left(\frac{(n-1)S^2}{0.02^2} \geq 67.505\right) = 0.05 = \alpha.$$

Dessutom för alla σ sådan att H_0 är sann så är

$$P\left(\frac{(n-1)S^2}{0.02^2} \geq 67.505\right) \leq \alpha.$$

Vi förkastar H_0 till förmån för H_1 om vi observerar ett utfall $(n-1)s^2/0.02^2$ större än 67.505. Här:

$$\frac{(n-1)s^2}{0.02^2} = 55.125$$

så H_0 förkastas ej på nivå $\alpha = 0.05$.

Kommentar: Då H_0 är sann så är

$$P\left(\frac{(n-1)S^2}{0.02^2} > 55.125\right) \leq 0.28701$$

så den lägsta signifikansnivå som H_0 skulle förkastas för är 0.28701, dvs testets p-värde.

14.6 Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(m, \sigma)$-fördelade stokastiska variabler beskrivandes mätningar av kvalitétsmåttet. Vi vill testa $H_0 : m = m_0$ mot $H_1 : m > m_0$ på nivå α. Vi förkastar H_0 till förmån för H_1 om $\overline{X}$ är stor, dvs om

$$Z = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}} > \lambda_\alpha$$

där Z är $\mathrm{N}(0, 1)$ om H_0 är sann. Alltså vi förkastar H_0 om $Z > \lambda_\alpha$ eller

$$\overline{X} > m_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}.$$

Styrkan bestäms genom

$$\begin{aligned}
\beta(m) &= P\left(\text{förkasta } H_0\right) = P\left(\frac{\overline{X} - m_0}{\sigma/\sqrt{n}} > \lambda_\alpha\right) = P\left(\frac{\overline{X} - m}{\sigma/\sqrt{n}} > \lambda_\alpha + \frac{m_0 - m}{\sigma/\sqrt{n}}\right) \\
&= 1 - \Phi\left(\lambda_\alpha + \frac{m_0 - m}{\sigma/\sqrt{n}}\right) = \Phi\left(-\lambda_\alpha + \frac{m - m_0}{\sigma/\sqrt{n}}\right).
\end{aligned}$$

Med styrkan $1 - \beta$ skall

$$\beta = 1 - \beta(m_1) = 1 - \Phi\left(-\lambda_\alpha + \frac{m_1 - m_0}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\lambda_\beta\right)$$

så

$$\lambda_\beta = -\lambda_\alpha + \frac{m_1 - m_0}{\sigma/\sqrt{n}}$$

eller

$$n = \left(\frac{(\lambda_\beta + \lambda_\alpha)\sigma}{m_1 - m_0}\right)^2.$$

Med $\alpha = 0.05$ och $1 - \beta = 0.90$ fås $\lambda_{0.05} = 1.6449$ och $\lambda_{0.10} = 1.2816$ så

$$n|_{m_1=1} = \left(\frac{(1.2816 + 1.6449) \cdot 1}{1 - 0}\right)^2 = 8.5638$$

dvs 9 observationer eller fler behövs för att nå den önskade styrkan. Vidare,

$$n|_{m_1=0.5} = \left(\frac{(1.2816 + 1.6449) \cdot 1}{0.5 - 0}\right)^2 = 34.255$$

(dvs 35 eller fler) och

$$n|_{m_1=0.1} = \left(\frac{(1.2816 + 1.6449) \cdot 1}{0.1 - 0}\right)^2 = 856.38$$

(dvs 857 eller fler).

14.7 Låt $X_1, \ldots, X_n$ beskriva alkoholhalten i flaskor öl från bryggeriet. Modell: $X_1, \ldots, X_n$ är oberoende och $\mathrm{N}(m, \sigma)$-fördelade där $\sigma = 0.001$. Vi vill testa:

$$H_0 : m = 0.03 = m_0 \quad \text{mot} \quad H_1 : m < m_0$$

på nivå $\alpha = 0.01$. Signifikansnivån begränsar sannolikheten för att förkasta en korrekt nollhypotes, dvs. här att dra slutsatsen att en flaska har liten alkoholhalt fast den egentligen är hög.

Vi skattar m med $\overline{X}$ och förkastar H_0 till förmån för H_1 om vi observerar ett litet värde på $\overline{X}$, alternativt, ett litet värde på

$$\frac{\overline{X} - m_0}{\sigma/\sqrt{n}}$$

som under H_0 är approximativt $\mathrm{N}(0,1)$-fördelad. Om H_0 är sann så är

$$P\left(\frac{\overline{X} - m_0}{\sigma/\sqrt{n}} \leq -\lambda_\alpha\right) = \alpha$$

eller, med sannolikhet α är

$$\overline{X} \leq m_0 - \lambda_\alpha \frac{\sigma}{n}.$$

Med $n = 10$ flaskor och signifikansnivå $\alpha = 0.01$ är $\lambda_\alpha = 2.3263$ så vi förkastar H_0 om

$$\overline{X} \leq m_0 - \lambda_\alpha \frac{\sigma}{n} = 0.03 - 2.3263 \frac{0.001}{\sqrt{10}} = 0.0293.$$

Vi observerar utfallet $\overline{x} = 0.0298$ och förkastar inte H_0 på nivå 1%.

14.8 Låt X beskriva resultatet av en smältpunktsmätning. Modell:

$$X = \text{smältpunkt} + \text{mätfel} = m + \epsilon$$

där ϵ är $\mathrm{N}(0, \sigma)$, $\sigma = 2.3$. Då är X $\mathrm{N}(m, \sigma)$. Låt $X_1, \ldots, X_n$ beskriva resultaten av n oberoende smältpunktsmätningar med observerade värden $x_1, \ldots, x_n$. Vi vill testa:

$$H_0 : m = m_0 = 1050^\circ \quad \text{mot} \quad H_1 : m \neq m_0$$

på nivå $\alpha = 0.05$. Vi förkastar H_0 för stora värden på $|\overline{X} - m_0|$, alternativt stora värden på $|Z|$ där

$$Z = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}}$$

som då H_0 är sann är N(0,1)-fördelad. Alltså, om H_0 är sann så är $P\left(|Z| > \lambda_{\alpha/2}\right) = \alpha$. 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

$$z = \frac{\overline{x} - m_0}{\sigma/\sqrt{n}} = \frac{1050.9 - 1050}{2.3/\sqrt{10}} = 1.2649$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

Kommentar: Då H_0 är sann så är

$$P\left(|Z| > 1.2649\right) = 2(1 - \Phi\left(1.2649\right)) = 0.2059$$

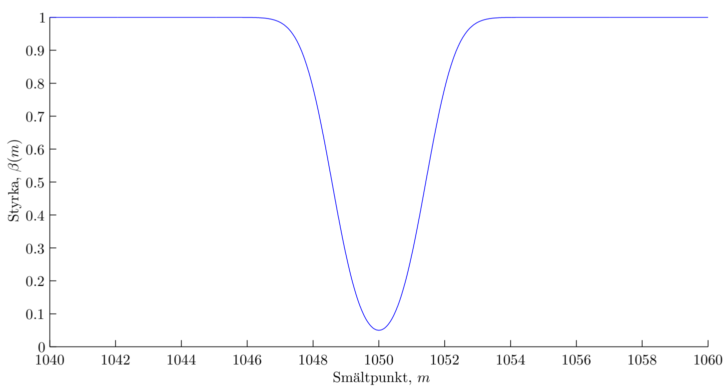
så den lägsta signifikansnivå vi skulle förkasta H_0 på är 0.2059, dvs testets p-värde.

Styrkan kan bestämmas som

$$\begin{aligned}
\beta(m) &= P\left(\text{förkasta } H_0\right) = P\left(\left|\frac{\overline{X} - m_0}{\sigma/\sqrt{n}}\right| > \lambda_{\alpha/2}\right) = 1 - P\left(-\lambda_{\alpha/2} \le \frac{\overline{X} - m_0}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2}\right) \\
&= 1 - P\left(-\lambda_{\alpha/2} + \frac{m_0 - m}{\sigma/\sqrt{n}} \le \frac{\overline{X} - m}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2} + \frac{m_0 - m}{\sigma/\sqrt{n}}\right) \\
&= 1 - \left(\Phi\left(\lambda_{\alpha/2} + \frac{m_0 - m}{\sigma/\sqrt{n}}\right) - \Phi\left(-\lambda_{\alpha/2} + \frac{m_0 - m}{\sigma/\sqrt{n}}\right)\right)
\end{aligned}$$

Vi beräknar

$$\beta(1051) = 0.27968 \qquad \beta(1053) = 0.9848.$$



Styrkefunktionen $\beta(m)$ som visar sannolikheten att testet förkastar H_0 för olika värden på värden på väntevärdet m.

Att inte förkasta H_0 betyder bara att H_0 inte är orimlig, inte att H_0 är sann. Vi kan inte påvisa att smältpunkten *är* 1050°, bara att det inte är oförenligt med våra mätresultat.

14.9 Låt $X_1, \ldots, X_n$ beskriva den uppmätta halten av ett ämne i en råvara. Modell: $X_1, \ldots, X_n$ är oberoende och N(m, σ)-fördelade där $\sigma = 0.2$. Vi vill testa:

$$H_0 : m = 5.5 = m_0 \quad \text{mot} \quad H_1 : m \ne m_0$$

på nivå $\alpha = 0.10$. Vi skattar m med $\overline{X}$ och förkastar H_0 till förmån för H_1 om vi observerar ett stort värde på $|\overline{X} - m_0|$, alternativt, ett stort värde på

$$\frac{|\overline{X} - m_0|}{\sigma/\sqrt{n}}$$

som under H_0 är approximativt beloppet av en N(0,1)-fördelad stokastisk variabel. Om H_0 är sann så är

$$P\left(\frac{|\overline{X} - m_0|}{\sigma/\sqrt{n}} \ge \lambda_{\alpha/2}\right) = \alpha$$

eller, med sannolikhet α är

$$|\overline{X} - m_0| \geq \lambda_{\alpha/2}\frac{\sigma}{n}.$$

Med $n = 6$ mätningar och signifikansnivå $\alpha = 0.10$ är $\lambda_{\alpha/2} = 1.6449$ så vi förkastar H_0 om

$$|\overline{X} \leq m_0| \geq \lambda_{\alpha/2}\frac{\sigma}{n} = 1.6449\,\frac{0.2}{\sqrt{6}} = 0.1343.$$

Med utfallet $\overline{x} = 5.575$ är $|\overline{x} - 5.5| = 0.075 < 0.1343$ så H_0 förkastas ej på nivå 10%. Med utfallet $\overline{x} = 5.675$ är $|\overline{x} - 5.5| = 0.175 > 0.1343$ så H_0 förkastas på nivå 10%.

Alternativt kan man först ta fram ett tvåsidigt symmetriskt konfidensintervall för m. Notera att vi inte förkastar H_0 för utfall

$$|\overline{x} - m_0| \leq \lambda_{\alpha/2}\frac{\sigma}{n}$$

dvs, för m_0 som uppfyller

$$\overline{x} - \lambda_{\alpha/2}\frac{\sigma}{n} \leq m_0 \leq \overline{x} + \lambda_{\alpha/2}\frac{\sigma}{n}$$

Alltså, konfidensintervallet för m, med konfidensgrad $1 - \alpha$,

$$m \in \overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

består precis av de värden på m_0 som vi inte förkastar H_0 för. Med observerat värde $\overline{x} = 5.575$ fås intervallet

$$m \in 5.575 \pm 0.1343 = [5.4407,\ 5.7093] \qquad (90\%)$$

och vi ser att värdet $m_0 = 5.5$ ingår i intervallet och således inte kan förkastas (på nivå 10%). Med observerat värde $\overline{x} = 5.675$ fås intervallet

$$m \in 5.675 \pm 0.1343 = [5.5407,\ 5.8093] \qquad (90\%)$$

och vi ser att värdet $m_0 = 5.5$ inte ingår i intervallet och således kan förkastas (på nivå 10%).

14.10 Låt $X_1, \ldots, X_n$ beskriva livslängderna för n stycken av tillverkare A:s glödlampor med observerade utfall $x_1, \ldots, x_n$. Mätningar ger värdena [i timmar]:

198	213	207	197	192
214	201	199	209	212

som sammanfattas med

$$\overline{x} = 204.2 \qquad s_x = 7.7574.$$

Modell: $X_1, \ldots, X_n$ är oberoende, likafördelade med $E(X_i) = m$ och $V(X_i) = \sigma^2$. Tillverkare A vill utföra testet

$$H_0 : m \leq 200 \quad \text{mot} \quad H_1 : m > 200$$

på nivå α, medan B vill utföra testet

$$H_0 : m \geq 200 \quad \text{mot} \quad H_1 : m < 200.$$

Med skattningen $\overline{x} = 204.2$ så kommer B inte att förkasta H_0 på någon vettig signifikansnivå.

Person A kommer att förkasta H_0 då $\overline{X}$ är mycket större än 200, alternativt då

$$T = \frac{\overline{X} - 200}{S/\sqrt{n}}$$

är stor. Med normalfördelningsantagande på $X_1, \ldots, X_n$ så för $m = 200$ (dvs H_0 är sann) är T $t(n-1)$-fördelad och $P(T > t_\alpha) = P(T > 1.8331) = 0.05 = \alpha$. För alla m sådana att H_0 är sann så är

$$P(T > 1.8331) \leq \alpha$$

så vi förkastar H_0 till förmån för H_1 då vi observerar ett utfall $T > 1.8331$. Vi observerar utfallet

$$t = \frac{\overline{x} - 200}{s_x/\sqrt{n}} = \frac{204.2 - 200}{7.7574/\sqrt{10}} = 1.7121 < 1.8331$$

så vi förkastar inte H_0 på nivå $\alpha = 0.05$. På nivå $\alpha = 0.10$ fås $t_\alpha = 1.3830$ så H_0 förkastas på nivå 0.10.

14.11 Låt $X_1, \ldots, X_n$ beskriva draghållfastheten vid n stycken B-dragprov med observerade utfall $x_1, \ldots, x_n$. Modell: $X_1, \ldots, X_n$ är oberoende och $\mathrm{N}(m, \sigma)$.

a) Vi vill utföra testet

$$H_0 : m \leq 48.0 \quad \text{mot} \quad H_1 : m > 48.0$$

på nivå α. Vi förkastar H_0 om $\overline{X}$ är mycket större än 48.0, alternativt då

$$T = \frac{\overline{X} - 48.0}{S/\sqrt{n}}$$

är stor. Då $m = 48.0$ så är T en $t(n-1)$-fördelad stokastisk variabel och $P(T > 1.3830) = \alpha = 0.10$. För alla m sådana att H_0 är sann gäller att $P(T > 1.3830) \leq \alpha$. Vi förkastar H_0 till förmån för H_1 då $T > 1.3830$. Vi observerar utfallet

$$t = \frac{\overline{x} - 48.0}{s/\sqrt{n}} = \frac{48.3 - 48.0}{1.5/\sqrt{20}} = 0.89443 < 1.3830$$

så vi förkastar inte H_0 på nivå 0.10. Med utfallen $\overline{x} = 49.7$ och $s = 1.7$ fås

$$t = \frac{\overline{x} - 48.0}{s/\sqrt{n}} = 4.4721 > 4.2968 = t_{0.001}$$

så H_0 förkastas på risknivå 0.001.

b) Om vi tror att B är bättre än A så byter vi om inte B är påvisbart sämre, dvs ifall vi kan förkasta när vi testar

$$H_0 : m \geq 48.0 \quad \text{mot} \quad H_1 : m < 48.0$$

på nivå α. Med skattningen $\overline{x} > 48.0$ kommer H_0 inte att förkastas på några rimliga signifikansnivåer.

14.12 Låt X beskriva resultatet av en smältpunktsmätning. Modell:

$$X = \text{smältpunkt} + \text{mätfel} = m + \epsilon$$

där ϵ är $\mathrm{N}(0, \sigma)$. Då är X $\mathrm{N}(m, \sigma)$. Låt $X_1, \ldots, X_n$ beskriva resultaten av n oberoende smältpunktsmätningar med observerade värden $x_1, \ldots, x_n$. Vi vill testa:

$$H_0 : m = m_0 = 1050^\circ \quad \text{mot} \quad H_1 : m \neq m_0$$

på nivå $\alpha = 0.05$. Vi förkastar H_0 för stora värden på $|\overline{X} - m_0|$, alternativt stora värden på $|T|$ där

$$T = \frac{\overline{X} - m_0}{S/\sqrt{n}}$$

som då H_0 är sann är $t(n-1)$-fördelad. Alltså, om H_0 är sann så är $P\left(|T| > t_{\alpha/2}\right) = \alpha$. Med $\alpha = 0.05$ är $t_{\alpha/2} = 2.2622$ och vi förkastar H_0 till förmån för H_1 om vi observerar ett utfall $|t| > 2.2622$. Här fås

$$t = \frac{\overline{x} - m_0}{s/\sqrt{n}} = \frac{1050.9 - 1050}{2.028/\sqrt{10}} = 1.4345$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

14.13 Låt $X_1, \ldots, X_n$ och $Y_1, \ldots, Y_n$ beskriva förslitningarna av däcktyp A resp. B på bil $1, 2, \ldots, n$. Modell: förslitningsskillnaderna

$$W_i = X_i - Y_i, \quad i = 1, \ldots, n$$

är oberoende och $\mathrm{N}(m, \sigma)$-fördelade stokastiska variabler. Vi vill testa hypotesen

$$H_0 : m = m_0 \quad \text{mot} \quad H_1 : m \neq m_0$$

på nivå $\alpha = 0.05$ där $m_0 = 0$. Från uppgift 13.24 fås konfidensintervallet för den genomsnittliga skillnaden i däckförslitning blir

$$m \in \overline{w} \pm t_{\alpha/2} \frac{s_w}{\sqrt{n}} = 0.10 \pm 0.1963 = [-0.0963,\ 0.2963] \qquad (95\%).$$

Konfidensintervallet består precis av de värden på m_0 som vi *inte* förkastar H_0 för. Eftersom värdet $m_0 = 0$ ingår i intervallet förkastar vi inte H_0 på nivå 5%.

14.14 Låt $X_1, \ldots, X_n$ och $Y_1, \ldots, Y_r$ beskriva bestämningar av molekylvikter på molekyler av typ A resp. typ B. Modell: alla stokastiska variabler är oberoende och

$$X = \text{molekylvikt typ A} + \text{mätfel} = m_x + \epsilon$$

där ϵ är $\mathrm{N}(0, \sigma)$. Då är X normalfördelad $\mathrm{N}(m, \sigma)$. På samma sätt är

$$Y = \text{molekylvikt typ B} + \text{mätfel} = m_y + \epsilon,$$

dvs Y är $\mathrm{N}(m_y, \sigma)$. Vi vill testa

$$H_0 : m_x = m_y \quad \text{mot} \quad H_1 : m_x \neq m_y$$

på nivå α. Hypoteserna formuleras istället

$$H_0 : m_x - m_y = 0 \quad \text{mot} \quad H_1 : m_x - m_y \neq 0.$$

Vi förkastar H_0 till förmån för H_1 om $|\overline{X} - \overline{Y}|$ är stor, alternativt då $|T|$ är stor där

$$T = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{r}}}$$

och

$$S = \sqrt{\frac{(n-1)S_x^2 + (r-1)S_y^2}{n + r - 2}}.$$

Då H_0 är sann så är T $t(n + r - 2) = t(12)$-fördelad och $P\left(|T| > t_{\alpha/2}\right) = P\left(|T| > 2.1788\right) = \alpha = 0.05$. Vi förkastar H_0 om utfallet på $|T|$ är större än $t_{\alpha/2}$.

Med data

$$n = 6,\ \overline{x} = 174.27,\ s_x = 0.062423 \quad r = 8,\ \overline{y} = 174.28,\ s_y = 0.091486$$

fås

$$s = \sqrt{\frac{(n-1)s_x^2 + (r-1)s_y^2}{n + r - 2}} = 0.080659$$

och

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{1}{n} + \frac{1}{r}}} = -0.18174.$$

Eftersom $|t| < t_{\alpha/2}$ förkastar vi inte H_0 på nivå $\alpha = 0.05$.

14.15 Låt $X_1, \ldots, X_n$ och $Y_1, \ldots, Y_r$ beskriva bestämningar av hållfastheten för lufttorkat resp. ugnstorkat trä. Modell: alla stokastiska variabler är oberoende och

$$X_i \text{ är } \mathrm{N}(m_x, \sigma) \qquad Y_i \text{ är } \mathrm{N}(m_y, \sigma).$$

Vi vill testa

$$H_0 : m_x - m_y \geq 10 \quad \text{mot} \quad H_1 : m_x - m_y < 10$$

på nivå α. Vi förkastar H_0 till förmån för H_1 för små värden på $\overline{X} - \overline{Y}$, alternativt då

$$T = \frac{(\overline{X} - \overline{Y}) - 10}{S\sqrt{\frac{1}{n} + \frac{1}{r}}}$$

är liten där

$$S = \sqrt{\frac{(n-1)S_x^2 + (r-1)S_y^2}{n+r-2}}.$$

Då $m_x - m_y = 10$ (dvs H_0 är sann) så är T en $t(n+r-2) = t(10+8-2)$-fördelad stokastisk variabel och $P(T < -t_\alpha) = P(T < -1.7459) = \alpha = 0.05$. För alla m sådana att H_0 är sann så är $P(T < -t_\alpha) \leq \alpha$. Vi förkastar H_0 om utfallet på T är mindre än $-t_\alpha$.

Med data

$$n = 10,\ \overline{x} = 104.36,\ s_x = 1.9783 \quad r = 8,\ \overline{y} = 96.087,\ s_y = 2.2068$$

fås

$$s = \sqrt{\frac{(n-1)s_x^2 + (r-1)s_y^2}{n+r-2}} = 2.0814$$

och

$$t = \frac{(\overline{x} - \overline{y}) - 10}{s\sqrt{\frac{1}{n} + \frac{1}{r}}} = -1.7498$$

Eftersom $t < -t_\alpha$ förkastar vi H_0 på nivå $\alpha = 0.05$.

14.16 Låt $X_1, \ldots, X_r$ beskriva antalet observationer i kategori $1, \ldots, r$, dvs antalet marsvinsungar med färg motsvarande kategori i.

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	$87 = n$
Hypotes: p_i	9/16	3/16	4/16	1
Förväntat antal: np_i	48.938	16.312	21.750	87

Vi jämför de observerade antalen x_i med de förväntade antalen np_i med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

som under H_0 är approximativt $\chi^2(r-1)$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(3-1) = \chi^2(2)$-tabeller får man att $\chi^2_{0.05} = 5.99$. Alltså: förkasta H_0 om $Q > 5.99$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = 10.063$$

så vi förkastar H_0 på nivå 5%. (Vi förkastar även på 1%.)

14.17 Låt $X_1, \ldots, X_r$ beskriva antalet observationer i kategori $1, \ldots, r$, dvs antalet utlånade böcker under de olika veckodagarna. Låt n vara det totala antalet utlånade böcker,

$$n = \sum_{i=1}^{r} X_i = 623.$$

Låt p_i, $i = 1, \ldots, r$ vara sannolikheten att en observation hamnar i kategori i. Vi vill testa en hypotes om att utlåningen är densamma oavsett veckodag mot att den varierar mellan veckodagar. Formellt: vi vill testa

$$H_0 : p_1 = p_2 = \cdots = p_r \quad \text{mot} \quad H_1 : \text{inte } H_0$$

på signifikansnivå $\alpha = 0.05$. Vår hypotes är att $p_1 = \cdots = p_5 = 1/5$. Om hypotesen är sann är det förväntade antalet observationer i kategori i, $E(X_i) = np_i$.

	Kategori (veckodag)					
	1 (mån)	2 (tis)	3 (ons)	4 (tor)	5 (fre)	
Observerat antal: x_i	135	108	120	114	146	$623 = n$
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 jämför de observerade antalen x_i med de förväntade antalen np_i med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

som under H_0 är approximativt $\chi^2(r-1)$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(5-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.05} = 9.49$. Alltså: förkasta H_0 om $Q > 9.49$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = \frac{(135 - 124.6)^2}{124.6} + \cdots + \frac{(146 - 124.6)^2}{124.6} = 7.8266 < 9.49$$

så vi förkastar inte H_0 på nivå 5%.

På nivå $\alpha = 0.10$ får man ur $\chi^2(4)$-tabeller att $\chi^2_{0.10} = 7.78 < q$. Alltså förkastar vi H_0 på nivå 10%.

14.18 De $n = 200$ observationerna delas in i $r = 5$ kategorier efter antal personer i kö:

	Kategori (antal kunder)					
	1 (0 st)	2 (1 st)	3 (2 st)	4 (3 st)	5 ($\geq$ 4 st)	
Observerad frekvens: x_i	30	42	34	28	66	$200 = n$

Notera att den sista kategorin är bara implicit given i uppgiften.

Låt p_i, $i = 1, \ldots, r$ vara sannolikheten att en observation hamnar i kategori i enligt statistikerns modell. Om denna fördelning stämmer är det förväntade antalet observationer i kategori i, $E(X_i) = np_i$.

	Kategori (antal kunder)					
	1 (0 st)	2 (1 st)	3 (2 st)	4 (3 st)	5 ($\geq$ 4 st)	
Observerat frekvens: x_i	30	42	34	28	66	$200 = n$
Hypotes: p_i	0.12	0.18	0.22	0.16	0.32	1
Förväntat antal: np_i	24	36	44	32	64	200

Vi jämför de observerade antalen x_i med de förväntade antalen np_i med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

som då hypotesen om fördelningen stämmer är approximativt $\chi^2(r-1)$-fördelad. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(5-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.05} = 9.49$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 9.49$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = \frac{(30-24)^2}{24} + \cdots + \frac{(66-64)^2}{64} = 5.3352 < 9.49$$

så vi förkastar inte hypotesen på nivå 5%. Vi kan inte utesluta att fördelningen är den givna.

14.19 Låt $X_1, \ldots, X_r$ beskriva antalet observationer i kategori $1, \ldots, r$, dvs antalet klockor med timvisaren i det givna intervallet. Låt n vara det totala antalet klockor

$$n = \sum_{i=1}^{r} X_i = 480.$$

Låt p_i, $i = 1, \ldots, r$ vara sannolikheten att en observation hamnar i kategori i. Vi vill testa en hypotes om att klockornas timvisare är likformigt fördelade över kategorierna. Formellt: vi vill testa

$$H_0 : p_1 = p_2 = \cdots = p_r \quad \text{mot} \quad H_1 : \text{inte } H_0$$

på signifikansnivå $\alpha = 0.05$. Vår hypotes är att $p_1 = \cdots = p_5 = 1/12$. Om hypotesen är sann är det förväntade antalet observationer i kategori i, $E(X_i) = np_i$.

	Kategori (tidsintervall)												
	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11	11–12	
Observerat antal: x_i	33	44	41	47	47	40	32	37	40	39	32	48	$480 = n$
Hypotes: p_i	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1
Förväntat antal: np_i	40	40	40	40	40	40	40	40	40	40	40	40	480

Vi jämför de observerade antalen x_i med de förväntade antalen np_i med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

som under H_0 är approximativt $\chi^2(r-1)$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(12-1) = \chi^2(11)$-tabeller får man att $\chi^2_{0.05} = 19.68$. Alltså: förkasta H_0 om $Q > 19.68$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = \frac{(33-40)^2}{40} + \cdots + \frac{(48-40)^2}{40} = 9.15 < 19.68$$

så vi förkastar inte H_0 på nivå 5%.

14.20 De $n = 100$ observationerna på den förmodat ffg-fördelade stokastiska variabeln delas in i 9 kategorier:

	Kategori (värde)									
	1	2	3	4	5	6	7	8	≥ 9	
Observerad frekvens: x_i	42	23	10	11	8	2	3	1	0	$100 = n$

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, $E(X_i) = np_i$.

	Kategori (värde)									
	1	2	3	4	5	6	7	8	≥ 9	
Observerad frekvens: x_i	42	23	10	11	8	2	3	1	0	100
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:

	Kategori (värde)					
	1	2	3	4	≥ 5	
Observerad frekvens: x_i	42	23	10	11	14	$100 = n$
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 jämför de observerade antalen x_i med de förväntade antalen np_i med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

som då hypotesen om ffg-fördelningen stämmer är approximativt $\chi^2(r-1)$-fördelad. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(5-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.01} = 13.3$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 13.3$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = \frac{(42-50)^2}{50} + \cdots + \frac{(14-6.25)^2}{6.25} = 15.16 > 13.3$$

så vi förkastar hypotesen på nivå 1%. Vi utesluter att observationerna kommer från en ffg(1/2)-fördelning.

14.21 De $n = 82$ observationerna på den förmodat Poisson-fördelade stokastiska variabeln delas in i 8 kategorier beroende på dess värde:

	Kategori (värde)								
	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.

Låt p_i vara sannolikheten att en observation hamnar i kategori i enligt statistikerns modell. Om Poisson(m)-antagandet stämmer skattar vi m med

$$m^* = 0 \cdot \frac{14}{81} + 1 \cdot \frac{12}{81} + \cdots + 6 \cdot \frac{1}{81} + 7 \cdot \frac{0}{81} = 2.1111.$$

Med denna skattning skattas

$$p_i = \frac{m^i}{i!} e^{-m} \quad \text{med} \quad p_i^* = \frac{m^{*i}}{i!} e^{-m^*}$$

för $i < 7$ och $p_7^* = 1 - \sum_{i=0}^{6} p_i^*$. De (skattade) förväntade antalet observationer i kategori i är np_i^*.

	Kategori (värde)								
	0	1	2	3	4	5	6	≥ 7	
Observerad frekvens: x_i	14	12	25	16	10	3	1	0	81
Hypotes: p_i^*	0.1211	0.25566	0.26987	0.18991	0.10023	0.042319	0.01489	0.0060259	1
Förväntat antal: np_i^*	9.8094	20.709	21.859	15.382	8.1185	3.4278	1.2061	0.4881	81

Kategorier med ett lågt förväntat antal, $np_i^* < 5$, slås samman för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 6$ kategorier:

	Kategori (värde)						
	0	1	2	3	4	≥ 5	
Observerad frekvens: x_i	14	12	25	16	10	4	81
Hypotes: p_i^*	0.1211	0.25566	0.26987	0.18991	0.10023	0.063234	1
Förväntat antal: np_i^*	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 np_i^* med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i^*)^2}{np_i^*}$$

som då hypotesen om ffg-fördelningen stämmer är approximativt $\chi^2(r-1-1)$-fördelad, vi tappar en extra frihetsgrad för att vi skattade parametern m. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(6-1-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.05} = 9.4877$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 9.4877$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i^*)^2}{np_i^*} = \frac{(14-9.8094)^2}{9.8094} + \cdots + \frac{(4-5.122)^2}{5.122} = 6.6105 < 9.4877$$

så vi förkastar ej hypotesen på nivå 5%. Vi kan inte utesluta att observationerna kommer från en Poissonfördelning.

14.22 De $n = 378$ observationerna på den förmodat normalfördelade stokastiska variabeln delas in i 10 kategorier beroende på dess värde:

	Kategori (värde)										
	<1	$1-2$	$2-3$	$3-4$	$4-5$	$5-6$	$6-7$	$7-8$	$8-9$	>9	
Observerad frekvens: x_i	0	1	1	2	33	139	155	42	5	0	378

Låt p_i vara sannolikheten att en observation hamnar i kategori i. Om normalfördelningsantagandet stämmer skattas väntevärdet m och standardavvikelsen σ med

$$m^* = 1.5 \cdot \frac{1}{378} + 2.5 \cdot \frac{1}{378} + \cdots + 7.5 \cdot \frac{42}{378} + 8.5 \cdot \frac{5}{378} = 6.0556$$

och

$$\sigma^* = \sqrt{\frac{1}{378-1}\left(1 \cdot 1.5^2 + 1 \cdot 2.5^2 + \cdots + 5 \cdot 8.5^2 - 378 m^{*2}\right)} = 0.91746.$$

Låt Y vara normalfördelad, $\mathrm{N}(m^*, \sigma^*)$. Då kan p_i skattas med $p_i^* = P(Y \in \text{kategori } i)$. De (skattade) förväntade antalet observationer i kategori i är np_i^*.

	Kategori (värde)				
	<1	$1-2$	$2-3$	$3-4$	$4-5$
Observerad frekvens: x_i	0	1	1	2	33
Hypotes: p_i^*	$1.7903 \cdot 10^{-8}$	$4.9076 \cdot 10^{-6}$	0.0004286	0.012096	0.11244
Förväntat antal: np_i^*	$6.7672 \cdot 10^{-6}$	0.0018551	0.16201	4.5724	42.501

	Kategori (värde)					
	$5-6$	$6-7$	$7-8$	$8-9$	>9	
Observerad frekvens: x_i	139	155	42	5	0	378
Hypotes: p_i^*	0.35089	0.3725	0.13461	0.016364	0.00066519	1
Förväntat antal: np_i^*	132.64	140.8	50.884	6.1856	0.25144	378

Kategorier med ett lågt förväntat antal, $np_i^* < 5$, slås samman med andra för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 6$ kategorier:

	Kategori (värde)						
	$<4, >9$	$4-5$	$5-6$	$6-7$	$7-8$	$8-9$	
Observerad frekvens: x_i	4	33	139	155	42	5	378
Hypotes: p_i^*	0.013195	0.11244	0.35089	0.3725	0.13461	0.016364	1
Förväntat antal: np_i^*	4.9877	42.501	132.64	140.8	50.884	6.1856	378

Vi jämför de observerade antalen x_i med de skattade förväntade antalen np_i^* med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i^*)^2}{np_i^*}$$

som då hypotesen om ffg-fördelningen stämmer är approximativt $\chi^2(r-1-2)$-fördelad, vi tappar två extra frihetsgrader för att vi skattade parametrarna m och σ. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(6-1-2) = \chi^2(3)$-tabeller får man att $\chi^2_{0.05} = 7.8147$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 7.8147$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i^*)^2}{np_i^*} = \frac{(4-4.9877)^2}{4.9877} + \cdots + \frac{(5-6.1856)^2}{6.1856} = 5.834 < 7.8147$$

så vi förkastar ej hypotesen på nivå 5%. Vi kan inte utesluta att observationerna kommer från en normalfördelning.

14.23 De $n = 300$ observationerna på den förmodat Poisson-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	$300 = n$

Låt p_i vara sannolikheten att en observation hamnar i kategori i enligt statistikerns modell. Om Poisson(m)-antagandet stämmer skattar vi m med

$$m^* = 0 \cdot \frac{249}{300} + 1 \cdot \frac{42}{300} + 2 \cdot \frac{9}{300} = 0.2$$

Med denna fås skattningarna

$$p_0^* = \frac{m^{*0}}{0!}\mathrm{e}^{-m^*} \quad p_1^* = \frac{m^{*1}}{1!}\mathrm{e}^{-m^*} \quad p_2^* = \frac{m^{*2}}{2!}\mathrm{e}^{-m^*} \quad p_3^* = 1 - \sum_{i=0}^{2} p_i^*.$$

De (skattade) förväntade antalet observationer i kategori i är np_i^*.

	Kategori (värde)				
	0	1	2	≥ 3	
Observerad frekvens: x_i	249	42	9	0	$300 = n$
Hypotes: p_i^*	0.81873	0.16375	0.016375	0.0011485	1
Förväntat antal: np_i^*	163.75	32.749	3.2749	0.2297	300

Kategorier med ett lågt förväntat antal, np_i^*, 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^*	0.81873	0.16375	0.017523	1
Förväntat antal: np_i^*	163.75	32.749	3.5046	300

Vi jämför de observerade antalen x_i med de skattade förväntade antalen np_i^* med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i^*)^2}{np_i^*}$$

som då hypotesen om ffg-fördelningen stämmer är approximativt $\chi^2(r-1-1)$-fördelad, vi tappar en extra frihetsgrad för att vi skattade parametern m. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(3-1-1) = \chi^2(1)$-tabeller får man att $\chi^2_{0.05} = 3.8415$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 3.8415$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i^*)^2}{np_i^*} = \frac{(249-163.75)^2}{163.75} + \cdots + \frac{(9-3.5046)^2}{3.5046} = 55.617 > 3.8415$$

så vi förkastar hypotesen på nivå 5%. Vi kan utesluta att observationerna kommer från en Poissonfördelning.

14.24 Med $n = 30$ observationer på den förmodat $\exp(m)$-fördelade stokastiska variabeln skattas parametern (väntevärdet) med medelvärdet

$$m^* = \frac{1}{30}(0.60 + 1.80 + \cdots + 1.87) = 1.000$$

Låt Y vara exponentialfördelad med parameter m^*. Observationerna delas in i 6 kategorier motsvarande intervallen

$$[0,\ 0.1823) \quad [0.1823,\ 0.4055) \quad [0.4055,\ 0.6932) \quad [0.6932,\ 1.099) \quad [1.099,\ 1.792) \quad [1.792,\ \infty)$$

Då fås att $p_i^* = P(Y \in \text{kategori i}) = 1/6$ och det förväntade antalet i varje kategori $np_i^* = 5$. Antalet observationer i respektive kategori sammanfattas i:

	Kategori						
	1	2	3	4	5	6	
Observerad frekvens: x_i	3	6	3	6	6	6	30
Hypotes: p_i^*	1/6	1/6	1/6	1/6	1/6	1/6	1
Förväntat antal: np_i^*	5	5	5	5	5	5	30

Vi jämför de observerade antalen x_i med de skattade förväntade antalen np_i^* med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i^*)^2}{np_i^*}$$

som då hypotesen om ffg-fördelningen stämmer är approximativt $\chi^2(r-1-1)$-fördelad, vi tappar en extra frihetsgrad för att vi skattade parametern m. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(6-1-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.05} = 9.4877$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 9.4877$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i^*)^2}{np_i^*} = \frac{(3-5)^2}{5} + \cdots + \frac{(6-5)^2}{5} = 2.4 < 9.4877$$

så vi förkastar ej hypotesen på nivå 5%. Vi kan inte utesluta att observationerna kommer från en exponentialfördelning.

14.25 Låt X_{ij} beskriva antalet observationer i kategori j, $j = 1, \ldots, r$, i mätserie i, $i = 1, \ldots, s$, dvs antalet personer som fick en viktförändring motsvarande kategori j när de åt piller motsvarande mätserie i. Låt n_i vara det totala antalet personer med piller i.

$$n_i = \sum_{j=1}^{r} X_{ij}, \qquad n_1 = 100,\ n_2 = 100,$$

med $n = \sum_{i=1}^{s} n_i = n_1 + n_2 = 200$ som totalt antal personer.

	Kategori (*viktförändring*)			
x_{ij}	1 (*mindre*)	2 (*oför*)	3 (*ökad*)	
Serie 1 *Bantyl*:	61	24	15	$100 = n_1$
Serie 2 *sockerpiller*:	54	29	17	$100 = n_2$
Totalt:	115	53	32	$200 = n$

Vi vill testa en hypotes om att viktförändring beror på pillersort. Om pillret inte påverkar vikten, låt $p_j,\ j=1,\dots,r$ vara sannolikheten att en observation hamnar i kategori j, dvs att en person får en viktförändring motsvarande kategori j. Formellt, vi vill testa huruvida sannolikheten för att en observation i serie i hamnar i kategori j beror av i, dvs

$$H_0 : \text{Pillersort påverkar ej viktförändring} \quad \text{mot} \quad H_1 : \text{ej } H_0$$

skrivs

$$H_0 : \begin{array}{l} P\left(\text{observation i serie } i \text{ hamnar i kategori } j\right) = p_j, \\ j=1,\dots,r, \quad \text{oavsett } i \end{array} \qquad \text{mot} \quad H_1 : \text{ej } H_0$$

Vi skattar p_j med

$$p_j^* = \frac{1}{n}\sum_{i=1}^{s} X_{ij} \qquad p_1^* = \frac{115}{200},\ p_2^* = \frac{53}{200},\ p_3^* = \frac{32}{200}.$$

Om pillret ej påverkar viktförändringen är det förväntade antalet observationer i kategori j i serie i $E\left(X_{ij}\right) = n_i p_j$ som skattas med $n_i p_j^*$.

	Kategori (*viktförändring*)			
Skattat förväntat, $n_i p_j^*$	1 (*mindre*)	2 (*oför*)	3 (*ökad*)	
Serie 1 *Bantyl*:	57.5	26.5	16	$100 = n_1$
Serie 2 *sockerpiller*:	57.5	26.5	16	$100 = n_2$
Totalt:	115	53	32	$200 = n$

Vi jämför de observerade antalen x_{ij} med de skattade förväntade antalen $n_i p_j^*$ med teststatistikan

$$Q = \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\left(X_{ij} - n_i p_j^*\right)^2}{n_i p_j^*}$$

som under H_0 är approximativt $\chi^2((s-1)(r-1))$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2((2-1)(3-1)) = \chi^2(2)$-tabeller får man att $\chi^2_{0.10} = 4.61$. Alltså: förkasta H_0 om $Q > 4.61$. Vi observerar utfallet

$$q = \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\left(x_{ij} - n_i p_j^*\right)^2}{n_i p_j^*} = 0.21304 + 0.23585 + 0.0625 + 0.21304 + 0.23585 + 0.0625 = 1.0228$$

så vi förkastar inte H_0 på nivå 10%.

14.26 Låt X_{ij} beskriva antalet observationer i kategori j, $j=1,\dots,r$, i mätserie i, $i=1,\dots,s$.

	Kategori (svarsalternativ)					
x_{ij}	1 (A)	2 (B)	3 (C)	4 (D)	5 (E)	
Serie 1: (pojkar)	130	230	220	370	50	$1000 = n_1$
Serie 2: (flickor)	110	310	210	250	120	$1000 = n_2$
Totalt:	240	540	430	620	170	$2000 = n$

Vi vill testa en hypotes om fördelningen för svarsalternativen skiljer sig åt mellan pojkar och flickor. Om det inte föreligger någon skillnad skattas p_j, sannolikheten att en utvald person ger ett svarsalternativ motsvarande kategori j, med

$$p_j^* = \frac{1}{n}\sum_{i=1}^{s} X_{ij} \qquad p_1^* = \frac{240}{2000},\ p_2^* = \frac{540}{2000},\ p_3^* = \frac{430}{2000},\ p_4^* = \frac{620}{2000},\ p_5^* = \frac{170}{2000}.$$

Vi kan då skatta det förväntade antalet observationer i kategori j i serie i $E(X_{ij}) = n_i p_j$ med $n_i p_j^*$.

Skattat förväntat, $n_i p_j^*$	1 (A)	2 (B)	3 (C)	4 (D)	5 (E)	
	Kategori (svarsalternativ)					
Serie 1: (pojkar)	120	270	215	310	85	1000
Serie 2: (flickor)	120	270	215	310	85	1000
Totalt:	240	540	430	620	170	2000

Vi jämför de observerade antalen x_{ij} med de skattade förväntade antalen $n_i p_j^*$ med teststatistikan

$$Q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(X_{ij} - n_i p_j^*)^2}{n_i p_j^*}$$

som under H_0 är approximativt $\chi^2((s-1)(r-1))$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2((2-1)(5-1)) = \chi^2(4)$-tabeller får man att $\chi^2_{0.01} = 13.28$. Alltså: förkasta H_0 om $Q > 13.28$. Vi observerar utfallet

$$q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - n_i p_j^*)^2}{n_i p_j^*} = 65.8$$

så vi förkastar H_0 på nivå 1%. Det är skillnad på hur pojkar och flickor svarar.

14.27 Låt X_{ij} beskriva antalet observationer i kategori j, $j = 1, \ldots, r$, i mätserie i, $i = 1, \ldots, s$, dvs antalet studenter som fick betyg motsvarande kategori j när de hade närvaro enligt mätserie i. Låt n_i vara det totala antalet studenter i mätserie i.

$$n_i = \sum_{j=1}^{r} X_{ij}, \qquad n_1 = 115,\ n_2 = 85,$$

med $n = \sum_{i=1}^{s} n_i = n_1 + n_2 = 200$ som totalt antal studenter.

Närvaro, x_{ij}	1 (*u*)	2 (*3*)	3 (*≥ 4*)	
	Kategori (*betyg*)			
Serie 1: $\geq$ *50%*	15	65	35	$115 = n_1$
Serie 2: $<$ *50%*	21	52	12	$85 = n_2$
Totalt:	36	117	47	$200 = n$

Vi vill testa en hypotes om att uppnått betyg beror på närvaron under lektionerna. Om närvaron inte påverkar betyget, låt p_j, $j = 1, \ldots, r$ vara sannolikheten att en observation hamnar i kategori j, dvs att en student får betyg motsvarande kategori j. Formellt, vi vill testa huruvida sannolikheten för att en observation i serie i hamnar i kategori j beror av i, dvs

$$H_0 : \text{Närvaro påverkar ej betygsfördelningen} \quad \text{mot} \quad H_1 : \text{ej } H_0$$

skrivs

$$H_0 : \begin{array}{l} P(\text{observation i serie } i \text{ hamnar i kategori } j) = p_j, \\ j = 1, \ldots, r, \quad \text{oavsett } i \end{array} \qquad \text{mot} \quad H_1 : \text{ej } H_0$$

Vi skattar p_j med

$$p_j^* = \frac{1}{n} \sum_{i=1}^{s} X_{ij} \qquad p_1^* = \frac{36}{200},\ p_2^* = \frac{117}{200},\ p_3^* = \frac{47}{200}.$$

Om närvaro ej påverkar betyget är det förväntade antalet observationer i kategori j i serie i $E(X_{ij}) = n_i p_j$ som skattas med $n_i p_j^*$.

Skattat förväntat, $n_i p_j^*$	1 (*u*)	2 (*3*)	3 (*≥ 4*)	
	Kategori (*betyg*)			
Serie 1: $\geq$ *50%*	20.700	67.275	27.025	$115 = n_1$
Serie 2: $<$ *50%*	15.300	49.725	19.975	$85 = n_2$
Totalt:	36	117	47	$200 = n$

Vi jämför de observerade antalen x_{ij} med de skattade förväntade antalen $n_i p_j^*$ med teststatistikan

$$Q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(X_{ij} - n_i p_j^*)^2}{n_i p_j^*}$$

som under H_0 är approximativt $\chi^2((s-1)(r-1))$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2((2-1)(3-1)) = \chi^2(2)$-tabeller får man att $\chi^2_{0.01} = 9.21$. Alltså: förkasta H_0 om $Q > 9.21$. Vi observerar utfallet

$$q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - n_i p_j^*)^2}{n_i p_j^*} = 1.5696 + 0.0769 + 2.3534 + 2.1235 + 0.1041 + 3.1840 = 9.4115$$

så vi förkastar H_0 på nivå 1%. Närvaron påverkar betyget.

14.28 Låt X_{ij} beskriva antalet observationer i kategori (i, j), $j = 1, \ldots, r$, $i = 1, \ldots, s$.

x_{ij}		Säkerhetsbälte 1 (användes)	Säkerhetsbälte 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

$$p_i^* = \frac{n_i}{N} \quad p_1^* = \frac{244}{500},\ p_2^* = \frac{256}{500}.$$

Vi skattar fördelningen för användandet av säkerhetsbälte med

$$q_i^* = \frac{m_i}{N} \quad q_1^* = \frac{159}{500},\ q_2^* = \frac{341}{500}.$$

Om graden av personskada är oberoende av användandet av säkerhetsbälte så kan man skatta sannolikheten för en observation i kategori (i, j), $p_i \cdot q_j$, med $p_i^* \cdot q_j^*$ och det förväntade antalet observationer med

$$N\, p_i^* \cdot q_j^* = \frac{n_i\, m_j}{N}.$$

$N\, p_i^* \cdot q_j^*$		Säkerhetsbälte 1 (användes)	Säkerhetsbälte 2 (användes ej)	
Personskador	1 (lätta)	77.592	166.408	244
	2 (svåra)	81.408	174.592	256
		159	341	500

Vi jämför de observerade antalen x_{ij} med de skattade förväntade antalen $Np_i^* q_j^*$ med teststatistikan

$$Q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(X_{ij} - \frac{n_i m_j}{N})^2}{\frac{n_i m_j}{N}}$$

som under H_0 är approximativt $\chi^2((s-1)(r-1))$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2((2-1)(2-1)) = \chi^2(1)$-tabeller får man att $\chi^2_{0.01} = 6.63$. Alltså: förkasta H_0 om $Q > 6.63$. Vi observerar utfallet

$$q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - \frac{n_i m_j}{N})^2}{\frac{n_i m_j}{N}} = 20.2235$$

så vi förkastar H_0 på nivå 1%. Användandet av säkerhetsbälte och graden av personskada är inte oberoende.

15.1 Låt $Y(x)$ vara normalfördelad med parametrar

$$Y(x) \text{ är } \mathrm{N}\left(\alpha' + \beta x, \sigma\right).$$

Med mätningar i $x_1, \ldots, x_n$ kan vi parametrisera modellen

$$Y(x) \text{ är } \mathrm{N}\left(\alpha + \beta(x - \overline{x}), \sigma\right).$$

där $\alpha = \alpha' + \beta\overline{x}$. Vi har följande $n = 7$ observationer:

x_i :	1.0	2.0	3.0	4.0	5.0	6.0	7.0
y_i :	0.9	1.4	2.2	2.7	3.2	4.3	4.2

Data kan sammanfattas med

$$\overline{x} = 4.0 \quad \overline{y} = 2.7$$

$$S_{xx} = \sum (x_i - \overline{x})^2 = 28 \quad S_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y}) = 16.7 \quad S_{yy} = \sum (y_i - \overline{y})^2 = 10.24$$

Detta ger oss skattningarna

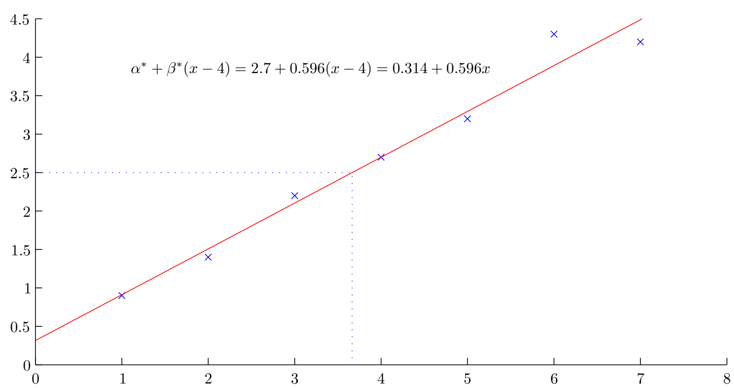
$$\alpha^* = \overline{y} = 2.7 \qquad \beta^* = \frac{S_{xy}}{S_{xx}} = 0.59643$$

och interceptet $\alpha' = \alpha - \beta x$ skattas med

$$(\alpha')^* = \alpha^* - \beta^*\overline{x} = 0.31429.$$

Vi söker x så att $\alpha' + \beta x = 2.5$, dvs $x = (2.5 - \alpha')/\beta$. En skattning ges av

$$\frac{2.5 - \alpha'^*}{\beta^*} = 3.6647.$$



Regressionslinjen $\alpha^* + \beta^*(x - 4) = 2.7 + 0.596(x - 4)$.

15.2 Låt $Y(x)$ vara normalfördelad med parametrar

$$Y(x) \text{ är } \mathrm{N}\left(\alpha' + \beta x, \sigma\right).$$

Med mätningar i $x_1, \ldots, x_n$ kan vi parametrisera modellen

$$Y(x) \text{ är } \mathrm{N}\left(\alpha + \beta(x - \overline{x}), \sigma\right).$$

där $\alpha = \alpha' + \beta\overline{x}$. Vi gör mätningar i punkter λ motsvarande $x = 1/\lambda^2$ 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

$$\overline{x} = 3.2138 \cdot 10^{-8} \quad \overline{y} = 25.033$$

$$S_{xx} = \sum(x_i - \overline{x})^2 = 8.0638 \cdot 10^{-17} \quad S_{xy} = \sum(x_i - \overline{x})(y_i - \overline{y}) = 6.8137 \cdot 10^{-8} \quad S_{yy} = \sum(y_i - \overline{y})^2 = 57.975.$$

Detta ger oss skattningarna

$$\alpha^* = \overline{y} = 25.033 \qquad \beta^* = \frac{S_{xy}}{S_{xx}} = 8.4497 \cdot 10^8$$

och interceptet $\alpha' = \alpha - \beta x$ skattas med

$$(\alpha')^* = \alpha^* - \beta^* \overline{x} = -2.1229.$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2}\left(S_{yy} - (\beta^*)^2 S_{xx}\right) = \frac{1}{n-2}\left(S_{yy} - \beta^* S_{xy}\right) = 0.40141$$

vilket ger $s = \sqrt{s^2} = 0.63357$.

15.3 Med observationer $y_1(x_1), \ldots, y_n(x_n)$, inför följande beteckningar:

$$S_{xx} = \sum(x_i - \overline{x})^2 \qquad S_{xy} = \sum(x_i - \overline{x})(y_i - \overline{y}) \qquad S_{yy} = \sum(y_i - \overline{y})^2.$$

Då är de oberoende skattningsvariablerna

$$\begin{array}{lll} \alpha^* & \text{är} & \mathrm{N}\left(\alpha, \sigma/\sqrt{n}\right) \\ \beta^* & \text{är} & \mathrm{N}\left(\beta, \sigma/\sqrt{S_{xx}}\right). \end{array}$$

En skattning av

$$m_0 = \alpha + \beta(x_0 - \overline{x})$$

ges av

$$m_0^* = \alpha^* + \beta^*(x_0 - \overline{x})$$

som är normalfördelad med väntevärde

$$E\left(m_0^*\right) = E\left(\alpha^* + \beta^*(x_0 - \overline{x})\right) = E\left(\alpha^*\right) + E\left(\beta^*\right)(x_0 - \overline{x}) = \alpha + \beta(x_0 - \overline{x}) = m_0$$

och varians

$$V\left(m_0^*\right) = V\left(\alpha^* + \beta^*(x_0 - \overline{x})\right) = V\left(\alpha^*\right) + V\left(\beta^*\right)(x_0 - \overline{x})^2 = \frac{\sigma^2}{n} + \frac{\sigma^2(x_0 - \overline{x})^2}{S_{xx}}$$

så

$$m_0^* \text{ är } \mathrm{N}\left(m_0, \sigma\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}\right).$$

Alltså är

$$\frac{m_0^* - m_0}{s\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}} \text{ är } t(n-2)$$

och ett konfidensintervall för m_0 med konfidensgrad $1 - p$ ges av

$$m_0 \in m_0^* \pm t_{p/2} s \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}.$$

15.4 Skattningsvariablerna

$$\begin{aligned} \alpha^* \quad &\text{är} \quad \mathrm{N}\left(\alpha, \sigma/\sqrt{n}\right) \\ \beta^* \quad &\text{är} \quad \mathrm{N}\left(\beta, \sigma/\sqrt{S_{xx}}\right). \end{aligned}$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2}\left(S_{yy} - (\beta^*)^2 S_{xx}\right) = \frac{1}{n-2}\left(S_{yy} - \beta^* S_{xy}\right) = 0.055929$$

vilket ger $s = \sqrt{s^2} = 0.23649$. Ur $t(n-2) = t(5)$-tabeller fås att $t_{0.025} = 2.5706$ så konfidensintervall med konfidensgrad 95% ges av

$$\begin{aligned} \alpha \quad &\in \quad \alpha^* \pm t_{0.025}\frac{s}{\sqrt{n}} = 2.7 \pm 0.22977 = [2.47,\ 2.93] \\ \beta \quad &\in \quad \beta^* \pm t_{0.025}\frac{s}{\sqrt{S_{xx}}} = 0.59643 \pm 0.11489 = [0.482,\ 0.711]. \end{aligned}$$

En skattning av

$$m_0 = \alpha + \beta(x_0 - \overline{x})$$

ges av

$$m_0^* = \alpha^* + \beta^*(x_0 - \overline{x})$$

som är normalfördelad med väntevärde och varians enligt:

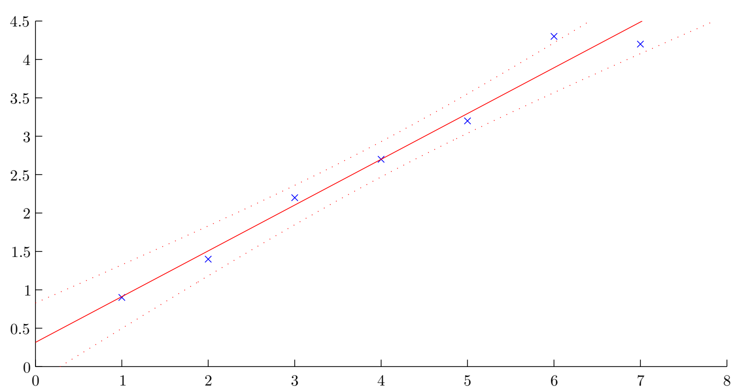
$$m_0^* \text{ är N}\left(m_0, \sigma\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}\right).$$

Ett konfidensintervall för m_0 med konfidensgrad 95% ges av

$$m_0 \in m_0^* \pm t_{0.025} s\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}.$$

Speciellt: med $x_0 = 0$ fås $m_0 = \alpha + (0 - \overline{x})\beta = \alpha'$ så är

$$\alpha' \in \alpha'^* \pm t_{0.025} s\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}} = 0.31429 \pm 0.51379 = [-0.1995,\ 0.82807] \quad (95\%).$$



Regressionslinjen från uppgift 15.1 med 95% konfidensintervallet för $\alpha + \beta x$ för olika värden på x inritad.

15.5 Fortsättning på uppgift 15.2.

Eftersom skattningsvariabeln $\alpha^* = \overline{Y}$ är $\mathrm{N}(\alpha, \sigma/\sqrt{n})$ får vi ett konfidensintervall av konfidensgrad $1-\gamma = 95\%$ som

$$\alpha \in \alpha^* \pm t_{\gamma/2}\frac{s}{\sqrt{n}} = 25.033 \pm 4.6478 \qquad (95\%).$$

Vidare, β^* är $\mathrm{N}(\beta, \sigma/\sqrt{S_{xx}})$ så ett ett konfidensintervall av konfidensgrad $1-\gamma = 95\%$ ges av

$$\beta \in \beta^* \pm t_{\gamma/2}\frac{s}{\sqrt{S_{xx}}} = 8.4497 \cdot 10^8 \pm 8.9648 \cdot 10^8 \qquad (95\%).$$

För att få ett konfidensintervall för $m(x) = \alpha + \beta(x - \overline{x})$ utnyttjar vi att

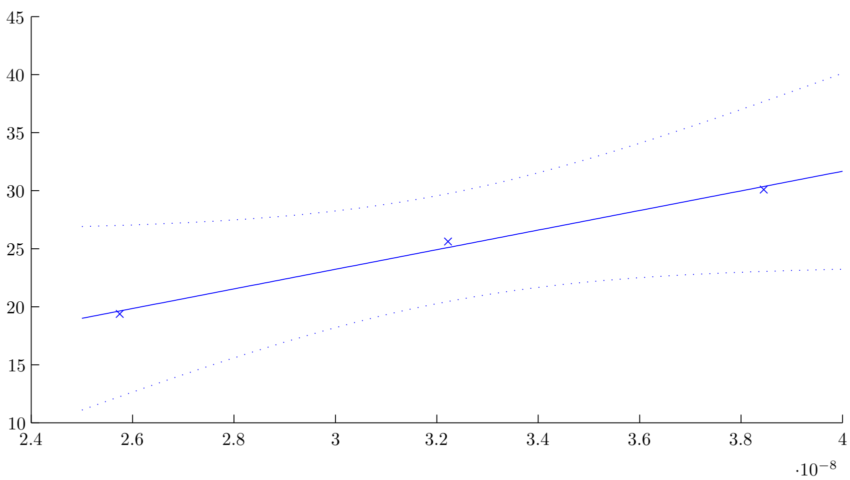
$$m^*(x) = \alpha^* + \beta^*(x - \overline{x}) \text{ är N}\left(m(x), \sigma\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}\right).$$

Alltså är

$$m(x) \in m^*(x) \pm t_{\gamma/2}s\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}.$$

För interceptet $\alpha' = m(0)$ får vi att

$$\alpha' \in (\alpha')^* \pm t_{\gamma/2}s\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}} = -2.1229 \pm 29.184 \qquad (95\%).$$



Observationerna (x_i, y_i) utprickade tillsammans med regressionslinjen $m^*(x) = \alpha'^* + \beta^* x$. Prickade linjerna markerar ändpunkterna av konfidensintervallet för $m(x)$ med konfidensgrad 95%.

B1421 En stokastisk variabel X kan anta värdena $S_X = \{0, 1, 2, 3\}$. Vi vill testa hypotesen

$$H_0 : X \text{ är Bin}(3, \frac{1}{4}) \qquad \text{mot} \qquad H_1 : X \text{ är ej Bin}(3, \frac{1}{4})$$

på nivå $\alpha = 0.05$. Om H_0 är sann så är sannolikheten att få en observation i kategori i

$$p_i = P(X = i) = \binom{3}{i} p^i (1-p)^{3-i}, \; i = 0, \ldots, 3$$

med $p = 1/4$. Med $n = 4096$ observationer låt X_i vara antalet observationer i kategori i.

i	0	1	2	3	
observerad frekvens, x_i	1764	1692	552	88	$n = 4096$
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 jämför de observerade antalen x_i med de förväntade antalen np_i med teststatistikan

$$Q = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

som då hypotesen om binomialfördelningen stämmer är approximativt $\chi^2(r-1)$-fördelad. Vi förkastar hypotesen för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2(4-1) = \chi^2(3)$-tabeller får man att $\chi^2_{0.01} = 11.35$. Alltså: förkasta hypotesen om den givna fördelningen om $Q > 13.3$. Vi observerar utfallet

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = 11.5 > 11.35$$

så vi förkastar hypotesen på nivå 1%. Vi utesluter att observationerna kommer från en $\mathrm{Bin}(3, 1/4)$-fördelning.

B1422 Låt X_{ij} beskriva antalet observationer i kategori j, $j = 1, \ldots, r$, i population i, $i = 1, \ldots, s$.

	Kategori (kön)		
x_{ij}	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:	267	233	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 så skattas p_j, sannolikheten att en utvald person har kön motsvarande kategori j, med

$$p_j^* = \frac{1}{n} \sum_{i=1}^{s} X_{ij} \qquad p_1^* = \frac{267}{500},\ p_2^* = \frac{233}{500}.$$

Vi kan då skatta det förväntade antalet observationer i kategori j i serie i $E(X_{ij}) = n_i p_j$ med $n_i p_j^*$.

	Kategori (kön)		
$n_i p_j^*$	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 $n_i p_j^*$ med teststatistikan

$$Q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(X_{ij} - n_i p_j^*)^2}{n_i p_j^*}$$

som under H_0 är approximativt $\chi^2((s-1)(r-1))$-fördelad. Vi förkastar H_0 för stora skillnader mellan observerat och förväntat, dvs. stora värden på Q. Ur $\chi^2((3-1)(2-1)) = \chi^2(2)$-tabeller får man att $\chi^2_{0.10} = 4.61$. Alltså: förkasta H_0 om $Q > 4.61$. Vi observerar utfallet

$$q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - n_i p_j^*)^2}{n_i p_j^*} = 3.7694 < 4.61$$

så vi förkastar inte H_0 på nivå 10%. Det är ingen på nivå 10% signifikant skillnad på andelarna män och kvinnor i populationerna.