Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/1915 SANNOLIKHETSTEORI OCH STATISTIK,
FREDAG 16 OKTOBER 2020 KL 8.00–13.00.

Examinator för SF1912: Mykola Shykula, 08-790 6644.
Examinator för SF1915: Björn-Olof Skytt, 08-790 8649.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen), miniräknare.

Tentamen består av två delar, benämnda del I och del II. Del I består av uppgifterna 1-12. På denna del skall endast svar anges, antingen i form av ett numeriskt värde med tre värdesiffrors noggrannhet eller i form av val av ett av de möjliga svarsalternativen. Svaren anges på svarsblanketten. Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter. Studenter som är godkända på datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift. Detta gäller på ordinarie tentamen och vid första omtentamen. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 8 poäng.
Del II består av uppgifterna 13-16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på eller får komplettera del I och poäng på del II krävs för högre betyg än E. På denna del skall resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar skall förklaras och definieras och numeriska svar skall anges med minst två värdesiffrors noggrannhet. Studenter som är godkända på datorlaborationen får 3 bonuspoäng på del II på ordinarie tentamenstillfället och det första omtentamenstillfället.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

Ett litet företag levererar paket till hushåll i en skärgårdskommun. 70% av paketen levereras med bil, 10% med cykel , och 20% med båt. Andelen skadade paket av de som levereras med bil är 2%. Om paketen levereras med cykel är andelen skadade paket 10%. Om leveransen sker med båt är andelen skadade paket 5%
Hur stor andel av alla levererade paket är skadade?

SVAR:..............

Uppgift 2

En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \begin{cases} 0, & x < 0 \\ 0.08x, & 0 \leq x \leq 5 \\ 0, & x > 5 \end{cases}$$

Bestäm $V(X)$.

A: 1.18

B: 1.39

C: 3.54

D: 12.50

Uppgift 3

Låt X,Y och Z vara oberoende stokastiska variabler sådana att $D(X) = 2$, $D(Y) = 5$, och $D(Z) = 3$.
Bestäm $D(3X - 2Z + Y - 3)$.

A: 5.06

B: 6.85

C: 8.06

D: 9.85

Uppgift 4

De stokastiska variablerna X och Y har den simultana sannolikhetsfunktionen

$p_{X,Y}(j,k)$	0	1	2
0	1/12	1/6	1/6
1	1/4	1/4	1/12

Således antar X värdena 0 och 1, medan Y antar värdena 0, 1, och 2.
Beräkna $P(X = 1|Y = 1)$.

A: 3/7

B: 1/4

C: 7/12

D: 3/5

Uppgift 5

Låt X vara en Normalfördelad stokastisk variabel med väntevärde $E(X) = -4$ och standardavvikelse $D(X) = 2$. D.v.s. $X \in N(-4, 2)$.
Bestäm a så att $P(X > a) = 0.05$.

A: -0.71

B: -0.08

C: 0.08

D: 0.71

Uppgift 6

På en verkstad arbetar 10 personer. Varje person använder sin skruvdragare 20% av arbetstiden. Händelserna att olika personer använder sin skruvdragare antas vara oberoende. Beräkna sannolikheten att minst 2 personer använder sin skruvdragare samtidigt.

A: 0.32

B: 0.59

C: 0.62

D: 0.68

Uppgift 7

Låt X_1, X_2 och X_3 vara tre oberoende stokastiska variabler sådana att $X_1 \in \mathrm{Po}\,(\mu)$, $X_2 \in \mathrm{Po}\,(3\mu)$, och $X_3 \in \mathrm{Po}\,(5\mu)$. Beräkna maximum-likelihood-skattningen av μ då $x_1 = 13$, $x_2 = 31$ och $x_3 = 55$.

SVAR:..............

Uppgift 8

Låt x vara en observation av $X \in Bin(n, p)$, där n är känd, men p är en okänd parameter. Minsta-kvadrat-skattningen av p ges då av

$$p^*_{obs} = \frac{x}{n},$$

dvs $p^* = \frac{X}{n}$.
Bestäm medelfelet för p^*, dvs bestäm $d(p^*)$.

A: $\frac{x(n-x)}{n}$

B: $\sqrt{\frac{x(n-x)}{n}}$

C: $\sqrt{\frac{x(n-x)}{n^2}}$

D: $\sqrt{\frac{x(n-x)}{n^3}}$

Uppgift 9

Antag att vi gör två stickprov $x_1, x_2, \ldots, x_n$ och $y_1, y_2, \ldots, y_m$ från två oberoende normalfördelade populationer $N(\mu_1, \sigma_1)$ respektive $N(\mu_2, \sigma_2)$. Stickprovet $x_1, x_2, \ldots, x_n$ gav medelvärdet $\overline{x}$ och standardavvikelsen s_x, medan stickprovet $y_1, y_2, \ldots, y_m$ gav medelvärdet $\overline{y}$ och standardavvikelsen s_y. Antag vidare att i populationen $\sigma_1 = \sigma_2$ och att de är okända. Vi testar $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$, på signifikans nivån 10%. Som testvariabel används: $t = \frac{\overline{x}-\overline{y}}{s\sqrt{\frac{1}{n}+\frac{1}{m}}}$, där $s^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2}$. Om $n = 12$ och $m = 10$, är beslutsregeln för att förkasta H_0:

A: $t > 1.33$

B: $|t| > 1.33$

C: $t > 1.72$

D: $|t| > 1.72$

Uppgift 10

Anta att $X \in Po(\mu)$ och låt H_0 vara att $\mu = 7$. Vi vill testa H_0 mot alternativet $\mu > 7$ och förkastar H_0 om vi observerar ett stort värde x på s.v. X. Antag att vi observerat $x = 13$. Bestäm testets P-värde.

A: 0.013

B: 0.027

C: 0.973

D: 0.987

Uppgift 11

Låt $X \in N(\mu, 1)$ och vi testar $H_0 : \mu = 0$. Låt x_1, x_2, x_3 och x_4 vara fyra oberoende mätningar på s.v. X. Vi beräknar medelvärdet $\bar{x} = \frac{\sum_{i=1}^{4} x_i}{4}$ och förkastar H_0 om $\bar{x} \geq 0.56$. Bestäm styrkan hos testet för alternativet $H_1 : \mu = 1$.

A: 0.67

B: 0.71

C: 0.81

D: 0.95

Uppgift 12

I ett avancerat växthus utförs ett experiment för att avgöra om mängden belysning påverkar hur mycket jordgubbar växer. Belysningen mäts med hjälp av ett belysningsindex och jordgubbarnas vikt mäts i gram. De första fyra erhållna observationerna följer nedan.

Belysningsindex	5	10	10	15
Jordgubbsvikt (g)	20	26	34	40

Det är rimligt att tro att det föreligger ett linjärt samband mellan variablerna. Utifrån datamaterialet ovan skattas en linjär regressionsmodell

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

där y_i = jordgubbsvikt (g) beror av x_i = belysningsindex och ε_i betecknar slumpmässiga fel, $i = 1, \ldots, 4$. Minsta-kvadrat-skattningarna av regressionskoefficienterna α och β blev $\alpha^*_{obs} = 10$ respektive $\beta^*_{obs} = 2$.

Vilket av de fyra svarsalternativen nedan motsvarar $I_\beta(0.9)$, dvs ett 90% konfidensintervall för den effekt som belysningindex har på jordgubbsvikten, om man vet att effekten i fråga är *signifikant* på 10% nivån.
Ledning: Inga beräkningar behövs för att lösa uppgiften.

A: $(0.35, 3.65)$

B: $(3.64, 16.36)$

C: $(-0.43, 4.43)$

D: $(-7.52, 27.52)$

Del II

Uppgift 13

I en dator beräknas det aritmetiska medelvärdet av 40 reella tal, som har avrundats till den närmaste tiondedelen. Som statistisk modell för avrundningsfelet för det i:te talet tas den stokastiska variabeln X_i, som är likformigt fördelad (= rektangelfördelad) i $[-0.05, 0.05]$, dvs. $X_i \in U(-0.05, 0.05)$ för $i = 1, \ldots, 40$. Avrundningsfelen antas vara oberoende från ett tal till annat. Beräkna approximativt sannolikheten för att det aritmetiska medelvärdet, såsom beräknat av datorn, skall överstiga det rätta värdet med mer än 0.01. Den införda approximationen skall motiveras. (10 p)

Uppgift 14

En person ska ta två bussar(en på morgonen och en på kvällen). De två väntetiderna X och Y är oberoende och $U(0, 10)$-fördelade.

a) Beräkna sannolikheten att den totala väntetiden blir högst 8. (2 p)

b) Bestäm fördelningsfunktionen $F(z)$ för den totala väntetiden $Z = X + Y$. (6 p)

c) Beräkna sannolikheten att den totala väntetiden blir högst 13. (2 p)

Ledning: (X, Y) är likformigt fördelad på kvadraten $K = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 10; 0 \leq y \leq 10\}$. Rita integrationsområdet.

Uppgift 15

Två olika analysmetoder för bestämning av nickelhalten i nickellegeringar jämfördes genom att 6 olika legeringar analyserades med båda metoderna. Följande värden erhölls:

Parti:	1	2	3	4	5	6
x:	20.66	14.35	23.14	15.42	17.9	16.54
y:	21.48	14.69	21.97	15.21	17.20	16.02

Låt oss anta att X_i:na kommer från en normalfördelning $N(\mu_i, \sigma_1)$ och att Y_i:na kommer från en normalfördelning $N(\mu_i + \Delta, \sigma_2)$. Avgör på risknivån 5% om det föreligger någon skillnad mellan analysmetoderna. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är. (10 p)

Uppgift 16

En balansvåg har mätfel som är en stokastisk variabel med väntevärde 0 och variansen σ^2. Två föremål med de okända vikterna θ_1 och θ_2 läggs först i samma vågskål och sedan i var sin vågskål. Härigenom får man en mätning x_1 av summan av vikterna och en mätning x_2 av skillnaden i vikt som är utfall av oberoende stokastiska variabler.

a) Bestäm minsta kvadrat-skattningen av θ_1 och θ_2 baserad på x_1 och x_2. (3 p)

b) Beräkna varianserna av dessa skattningar. Jämför dessa varianser med vad man skulle få om man vägde ett föremål i taget. Dvs motivera vilken metod att skatta som är effektivast. (2 p)

c) Undersök om minsta kvadrat-skattningarna av θ_1 och θ_2 är okorrelerade (dvs att de har kovarians=0). (5 p)

Kommentar: Ovanstående uppgift är ett specialfall av det s k Hotellings vägningsproblem. Den intressserade kan hemma fundera på hur man med en balansvåg lämpligen väger 4 föremål då man får göra totalt 4 vägningar!

Lycka till!

Avd. Matematisk statistik

LÖSNINGSFÖRSLAG TENTAMEN I SF1912/SF1915 SANNOLIKHETSTEORI OCH STATISTIK, FREDAG 16 OKTOBER 2020 KL 8.00–13.00.

Del I

1. 0.034

2. B

3. D

4. D

5. A

6. C

7. 11

8. D

9. D

10. B

11. C

12. A

Uppgift 1

Här har vi lagen om total sannolikhet. Låt A, B, C beteckna händelserna att en leveransen sker med bil, cykel respektive båt. Låt S beteckna händelsen att ett levererat paket är skadat. Då blir

$$P(S) = P(S|A)P(A) + P(S|B)P(B) + P(S|C)P(C) = 0.02 \cdot 0.70 + 0.10 \cdot 0.10 + 0.05 \cdot 0.20 = 0.034$$

Uppgift 2

$$E(X) = \int_0^5 x \cdot 0.08 x dx = \left[0.08\frac{x^3}{3}\right]_0^5 = \frac{10}{3}$$

$$E(X^2) = \int_0^5 x^2 \cdot 0.08 x dx = \left[0.08\frac{x^4}{4}\right]_0^5 = \frac{125}{10}$$

$$V(X) = E(X^2) - E^2(X) = \frac{125}{10} - \frac{100}{9} = 1.39$$

Uppgift 3

$V(3X - 2Z + Y - 3) =$[alla stokastiska variablerna är oberoende]$= V(3X) + V(-2Z) + V(Y) =$
$9V(X) + 4V(Z) + V(Y) = 9 \cdot 4 + 4 \cdot 9 + 25 = 97 \Rightarrow D(3X - 2Z + Y - 3) = \sqrt{97} = 9.85$

Uppgift 4

$$P(X = 1|Y = 1) = \frac{P(X = 1 \cap Y = 1)}{P(Y = 1)} = \frac{\frac{1}{4}}{\frac{1}{6} + \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{10}{24}} = \frac{3}{5}$$

Uppgift 5

$P(X > a) = 0.05.$
Gör om till $N(0, 1)$ när $E(X) = -4$ och $D(X) = 2$.

$$\Rightarrow P(\frac{X + 4}{2} > \frac{a + 4)}{2} = 0.05$$

$$\Rightarrow \frac{a + 4}{2} = \lambda_{0.05}$$

$\Rightarrow a = 2 \cdot \lambda_{0.05} - 4 = -0.71$

Uppgift 6

Låt X vara antalet personer som använder sin skruvdragare samtidigt. Då gäller att $X \in Bin(10, 0.2)$
$P(X \geq 2) = 1 - P(X \leq 1) =$[se tab 6]$1 - 0.38 = 0.62$

Uppgift 7

se § 9.1 i F.S.

$$L(\mu) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot p_{X_3}(x_3) = p_{X_1}(13) \cdot p_{X_2}(31) p_{X_3}(55) = \frac{(\mu)^{13}}{13!} e^{-\mu} \cdot \frac{(3\mu)^{31}}{31!} e^{-3\mu} \cdot \frac{(5\mu)^{55}}{55!} e^{-5\mu} =$$

$$= \frac{\mu^{13+31+55}}{13!31!55!} \cdot 3^{31} \cdot 5^{55} e^{(-1-3-5)\mu} = \frac{\mu^{99}}{13!31!55!} \cdot 3^{31} \cdot 5^{55} e^{-9\mu}$$

$$ln(L(\mu)) = 99 ln(\mu) + ln\left(\frac{3^{31} \cdot 5^{55}}{13!31!55!}\right) - 9\mu$$

$$\frac{d}{d\mu} ln(L(\mu)) = 0 \Rightarrow \frac{99}{\mu} - 9 = 0 \Rightarrow \mu = 11$$

Alltså är Maximum-Likelihood-skattningen av μ lika med 11. D.v.s. $\mu^*_{obs_{ML}} = 11$

Uppgift 8

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

$$D(p^*) = \sqrt{\frac{1}{n} p(1-p)} \Rightarrow D^*_{obs}(p^*) = \sqrt{\frac{1}{n} p^*_{obs}(1 - p^*_{obs})} = \sqrt{\frac{1}{n} \frac{x}{n} (1 - \frac{x}{n})} = \sqrt{\frac{1}{n^3} x(n-x)}$$

D.v.s. medelfelet för p^* är

$$d(p^*) = \sqrt{\frac{x(n-x)}{n^3}}$$

Uppgift 9

Här har vi hypotestest för två oberoende stickprov (från två oberoende populationer). Signifikansnivån är $\alpha = 10\%$. Eftersom $H_1 : \mu_1 \neq \mu_2$ är tvåsidigt, och $n = 12$, $m = 10$, har vi att beslutsregeln blir följande: förkasta H_0 om

$$|t| > t_{\frac{\alpha}{2}}(n + m - 2),$$

där $t_{\frac{\alpha}{2}}(n + m - 2) = t_{0.05}(20) = 1.72$, enligt Tabell 3. Dvs svarsaltenativ D är det korrekta.

Uppgift 10

p-värdet = P(förkasta H_0) om H_0 är sann och $x \geq 13 \Leftrightarrow P(X \geq 13)$ om $X \in Po(7)$

$P(X \geq 13) = 1 - P(X \leq 12) =$ [se tab 5] $= 1 - 0.973 = 0.027$

Uppgift 11

Styrkan hos testet= P(förkasta H_0) om H_1 är sann = P($\bar{X} \geq 0.56$) om $\mu = 1 \Rightarrow$ att styrkan hos testet blir

$$P(\bar{X} \geq 0.56) = P(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{0.56 - \mu}{\frac{\sigma}{\sqrt{n}}})$$

Vi vet dock att n=4, $\sigma = 1$, och att $\mu = 1$ enligt H_1.
Styrkan hos testet blir då

$$P(\frac{\bar{X} - 1}{\frac{1}{\sqrt{4}}} > \frac{0.56 - 1}{\frac{1}{\sqrt{4}}}) = P(\frac{\bar{X} - 1}{\frac{1}{\sqrt{4}}} > -0.88) = \Phi(0.88) = 0.8106$$

Uppgift 12

1) Konfidensintervallet I_β måste innehålla $\beta^*_{obs} = 2$, eftersom intervallet är symmetriskt runt β^*_{obs}.
2) Konfidensintervallet I_β ska *inte* innehålla noll, eftersom effekten som belysningsindex har på jordgubbsvikten är signifikant.
Det finns bara ett intervall (nämligen, i svarsalternativ A) som uppfyller de två kraven.

Del II

Uppgift 13

Om $X_i \in U(-0.05, 0.05)$, fås enligt formelsamlingen att

$$E(X_i) = 0, V(X_i) = \frac{(0.05 - (-0.05))^2}{12} = \frac{1}{1200}.$$

Avrundningsfelet i det aritmetiska medelvärdet är

$$\overline{X}_{40} = \frac{1}{40}(X_1 + \ldots + X_{40}).$$

Då gäller att

$$E\left(\overline{X}_{40}\right) = 0$$

och eftersom avrundningsfelen är oberoende

$$D\left(\overline{X}_{40}\right) = \frac{1}{\sqrt{40}} \cdot \frac{1}{\sqrt{1200}}.$$

Ifall $\overline{X}_{40} > 0.01$ kommer det av datorn beräknade medelvärdet att överstiga det rätta värdet med mer än 0.01. SÖKT är med andra ord $P\left(\overline{X}_{40} > 0.01\right)$. Den *centrala gränsvärdessatsen* innebär att $\overline{X}_{40}$ är, som en summa av oberoende likafördelade stokastiska variabler, approximativt normalfördelad med fördelningen

$$N\left(0, D\left(\overline{X}_{40}\right)\right) = N\left(0, \frac{1}{\sqrt{40}} \cdot \frac{1}{\sqrt{1200}}\right).$$

Detta ger att

$$P\left(\overline{X}_{40} > 0.01\right) = P\left(\frac{\overline{X}_{40} - 0}{D\left(\overline{X}_{40}\right)} > \frac{0.01 - 0}{D\left(\overline{X}_{40}\right)}\right) =$$

$$= 1 - P\left(\frac{\overline{X}_{40}}{D\left(\overline{X}_{40}\right)} \leq \frac{0.01}{D\left(\overline{X}_{40}\right)}\right) \approx 1 - \Phi\left(0.01 / \frac{1}{\sqrt{40}} \cdot \frac{1}{\sqrt{1200}}\right) =$$

$$= 1 - \Phi\left(0.01 \cdot \sqrt{40} \cdot \sqrt{1200}\right) \approx 1 - \Phi(2.19),$$

där $0.01 \cdot \sqrt{40} \cdot \sqrt{1200}$ avrundats till 2.19. En tabellslagning ger $\Phi(2.19) = 0.98574$. SVAR: $\underline{P\left(\overline{X}_{40} > 0.01\right) \approx 0.0143}$.

Uppgift 14

a) Tänk geometriskt shl. Utfallsrummet är $\Omega = K = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 10, 0 \leq y \leq 10\}$. Rita kvadraten K på (x, y)-plan. Vi söker $P(X+Y \leq 8)$. Låt $T_1 = \{(x, y) \in K : x+y \leq 8\} = \{(x, y) \in K : y \leq 8 - x\}$. Rita arean T_1: T_1 är en rätvinklig triangel. Enligt konceptet av den geometriska sannolikheten för den likformiga tvådimensionella fördelningen (se s.88 i Blom), har vi:

$$P(X + Y \leq 8) = |\text{geometrisk sannolikh}| = \frac{\text{area of } T_1}{\text{area of } \Omega} = \frac{\text{area of } T_1}{\text{area of K}} = \frac{8^2/2}{10^2} = \frac{64}{200} = 0.32$$

Svar: 0.32

c) Låt nu $T_2 = \{(x, y) \in K : x + y \leq 13\} = \{(x, y) \in K : y \leq 13 - x\}$. Låt komplementet vara $T_2^* = \{(x, y) \in K : x + y > 13\} = \{(x, y) \in K : y > 13 - x\}$. Notera att T_2^* är också en rätvinklig triangel (rita på (x, y)-plan). Med samma resonemang som i a) och genom att rita aren T_2 och T_2^*, har vi:

$$P(X + Y \leq 13) = P(T_2) = |\text{Komplement}| = 1 - P(T_2^*) = 1 - \frac{\text{area of } T_2^*}{\text{area of } K} = |\text{rita figur}|$$
$$= 1 - \frac{7^2/2}{10^2} = 1 - \frac{49}{200} = 0.755$$

Svar: 0.755

b) Fördelningsfunktionen är enligt definitionen: $F(z) = P(Z \leq z) = P(X + Y \leq z)$.
För $0 \leq z \leq 10$, på samma sätt som i a):

$$F(z) = P(X + Y \leq z) = |\text{rita figur}| = \frac{z^2/2}{10^2} = \frac{z^2}{200}.$$

För $10 \leq z \leq 20$, på samma sätt som i c):

$$F(z) = P(X + Y \leq z) = |\text{rita figur, komplement}| = 1 - \frac{\big(10 - (z - 10)\big)^2/2}{10^2} = 1 - \frac{(20 - z)^2}{200}.$$

Svar: Sammanfattningsvis,

$$F(z) = \begin{cases} 0\ , & z < 0, \\ \dfrac{z^2}{200}\ , & 0 \leq z < 10, \\ 1 - \dfrac{(20 - z)^2}{200}\ , & 10 \leq z < 20, \\ 1\ , & z \geq 20. \end{cases}$$

Uppgift 15

Här har vi stickprov i par. Nollhypotesen är H_0: $\Delta = 0$. Mothypotesen är H_1: $\Delta \neq 0$: Risknivån är 5% och vi kan anta att $Z_i = Y_i - X_i \in N(\Delta, \sigma_z)$. Vi bildar ett 95%-igt konf-int.

$$I_\Delta = \bar{z} \pm t_{\alpha/2}(5) \cdot \frac{s_z}{\sqrt{6}} = -0.24 \pm 2.57 \cdot \frac{0.72}{\sqrt{6}} = -0.24 \pm 0.76.$$

$0 \in I_\Delta$. D.v.s. vi kan inte påvisa någon signifikant skillnad på 5%-nivån.

Uppgift 16

x_1 är ett utfall av X_1 där $E(X_1) = \theta_1 + \theta_2$ och $V(X_1) = \sigma^2$
x_2 är ett utfall av X_2 där $E(X_2) = \theta_1 - \theta_2$ och $V(X_2) = \sigma^2$.
a) Minsta kvadratmetoden innebär att vi studerar

$$Q(\theta_1, \theta_2) = (x_1 - (\theta_1 + \theta_2))^2 + (x_2 - (\theta_1 - \theta_2))^2$$

och vi erhåller

$$\frac{\partial Q}{\partial \theta_1} = 2(x_1 - \theta_1 - \theta_2)(-1) + 2(x_2 - \theta_1 + \theta_2)(-1)$$

$$\frac{\partial Q}{\partial \theta_2} = 2(x_1 - \theta_1 - \theta_2)(-1) + 2(x_2 - \theta_1 + \theta_2)$$

och ekvationssystemet

$$\frac{\partial Q}{\partial \theta_1} = 0, \quad \frac{\partial Q}{\partial \theta_2} = 0$$

har lösningen

$$\theta_1^* = \frac{1}{2}(x_1 + x_2) \text{ och } \theta_2^* = \frac{1}{2}(x_1 - x_2).$$

b) Vi får

$$V(\theta_1^*) = V\left(\frac{1}{2}(X_1 + X_2)\right) = \frac{1}{4}(V(X_1) + V(X_2)) = \frac{\sigma^2}{2}$$

och på samma sätt blir $V(\theta_2^*) = \sigma^2/2$. Vid vägning av ett föremål i taget skulle vi få varianserna σ^2 för vardera av skatningarna av θ_1 och θ_2. Detta betyder att det är effektivare att väga summan och skillnaden än att väga föremålen ett i taget!!

c) Vi får eftersom $C(X, Y) = E(XY) - E(X)E(Y)$ att

$$C(\theta_1^*, \theta_2^*) = C\left(\frac{1}{2}(X_1 + X_2), \frac{1}{2}(X_1 - X_2)\right) =$$

$$= E\left(\frac{1}{4}(X_1 + X_2)(X_1 - X_2)\right) - E\left(\frac{1}{2}(X_1 + X_2)\right) E\left(\frac{1}{2}(X_1 - X_2)\right) =$$

$$= \frac{1}{4}E(X_1^2 - X_2^2) - \frac{1}{4}(\theta_1 + \theta_2 + \theta_1 - \theta_2)(\theta_1 + \theta_2 - (\theta_1 - \theta_2)) =$$

$$= \frac{1}{4}E(X_1^2) - \frac{1}{4}E(X_2^2) - \theta_1\theta_2 = \frac{1}{4}\left(E(X_1^2) - (\theta_1 + \theta_2)^2\right) - \frac{1}{4}\left(E(X_2^2) - (\theta_1 - \theta_2)^2\right) =$$

$$= \frac{1}{4}(V(X_1) - V(X_2)) = 0.$$

Förvånansvärt nog är alltså skattningarna θ_1^* och θ_2^* okorrelerade.

Svar till kommentaren: För 4 föremål kan man t ex väga
$\theta_1 + \theta_2 + \theta_3 + \theta_4$
$\theta_1 - \theta_2 + \theta_3 - \theta_4$
$\theta_1 + \theta_2 - \theta_3 - \theta_4$
$\theta_1 - \theta_2 - \theta_3 + \theta_4$
som ger skattningar med variansen $\sigma^2/4$ och som dessutom är okorrelerade att jämföra med variansen σ^2 om man väger ett föremål i taget!