Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1917/SF1918/SF1919 SANNOLIKHETSTEORI OCH STATISTIK, ONSDAG 8 JANUARI 2020 KL 8.00–13.00.

Examinator: Camilla Landén, 08-790 6197.

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. 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 vid 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 vid 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

För händelserna A och B gäller att $P(A^* \cap B^*) = 0.1$, $P(A^* \cup B^*) = 0.7$ och $P(B) = 0.6$. Bestäm $P(A)$.

A: 0.2

B: 0.4

C: 0.6

D: 0.8

Uppgift 2

En stokastisk variabel X har fördelningsfunktionen

$$F(x) = \begin{cases} 0 & \text{om } x < 11 - c, \\ \frac{1}{2c}(x - 11 + c) & \text{om } 11 - c \leq x \leq 11 + c, \\ 1 & \text{om } x > 11 + c, \end{cases}$$

där $c > 0$. Antag att $V(X) = 12$. Bestäm konstanten c.

Uppgift 3

Låt X och Y vara stokastiska variabler, sådana att $V(X) = 2$, $V(Y) = 5$, och korrelationskoefficienten för X och Y är $\rho(X, Y) = -\frac{2}{3}$.
Beräkna $V(3X - 4Y)$.

A: 47.4

B: 76.6

C: 102.4

D: 148.6

Uppgift 4

Det är känt att 70% av de som äter på det lokala konditoriet beställer kaffe. Antag att de följande 10 kunderna beställer oberoende av varandra. Vad är sannolikheten att minst hälften av dem beställer kaffe?

A: 65 %

B: 75 %

C: 85 %

D: 95 %

Uppgift 5

Låt X, Y och Z vara oberoende stokastiska variabler, sådana att $X \in Po(1)$, $Y \in Po(0.5)$ och $Z \in Po(0.9)$.
Bestäm $P(X + Y + Z = 2)$.

Uppgift 6

Som modell för att beskriva variationen av poäng på antagningsprovet till en viss utbildning kan man använda en normalfördelning med väntevärdet 500 och standardavvikelsen 100.
Vilken antagningsgräns ska man sätta om man vill att ungefär 10% av de med högst poäng ska komma in?

A: 576

B: 628

C: 656

D: 664

Uppgift 7

Låt x_i, $i = 1, 2, 3$ vara observationer av de oberoende stokastiska variablerna X_i, $i = 1, 2, 3$ vilka alla är gammafördelade $\Gamma(2, \lambda)$, dvs de har täthetsfunktion

$$f_{X_i}(x) = \begin{cases} \dfrac{\lambda^2}{1!} \cdot x e^{-\lambda x} & \text{om } x > 0, \\ 0 & \text{om } x \leq 0 \end{cases} \qquad i = 1, 2, 3.$$

för $\lambda > 0$. Detta medför att

$$E(X_i) = \frac{2}{\lambda} \quad \text{och} \quad V(X_i) = \frac{2}{\lambda^2} \quad i = 1, 2, 3.$$

Antag att följande värden har erhållits:

$$x_1 = 1.4075, \qquad x_2 = 0.7142, \qquad x_3 = 0.4004.$$

Bestäm MK-skattningen av λ baserat på de givna värdena.

A: 1.681

B: 2.379

C: 0.420

D: 0.595

Uppgift 8

Kolesterolhalten i blodet mäts hos fem personer före och efter en period med förändrad kost och motion. Antag att mätvärdena är normalfördelade.

	1	2	3	4	5
Före	262	272	284	298	294
Efter	252	262	278	282	278

Ange med 95% konfidensintervall vilken genomsnittlig förändring av kolesterolhalt som kan påvisas.

A: $[6.21,\ 16.99]$

B: $[7.47,\ 15.73]$

C: $[-8.7870,\ 31.9870]$

D: $[-5.7277,\ 28.9277]$

Uppgift 9

Antag att $x_1, \ldots, x_n$ utgör ett stort stickprov från en fördelning med väntevärde μ och standardavvikelse σ, där både μ och σ är okända. Tyko önskar testa nollhypotesen $H_0 : \mu = 5$ mot $H_1 : \mu < 5$ med hjälp av ett lämpligt konfidensintervall för μ. Låt

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{och} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Vilket av nedanstående konfidensintervall för μ kan användas för att genomföra hypotestestet på approximativ signifikansnivå α?

A: $I_\mu = \left(\overline{x} - \frac{\sigma}{\sqrt{n}} \cdot \lambda_{\alpha/2},\ \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot \lambda_{\alpha/2}\right)$

B: $I_\mu = \left(\overline{x} - \frac{s}{\sqrt{n}} \cdot \lambda_{\alpha/2},\ \overline{x} + \frac{s}{\sqrt{n}} \cdot \lambda_{\alpha/2}\right)$

C: $I_\mu = \left(\overline{x} - \frac{\sigma}{\sqrt{n}} \cdot \lambda_{\alpha}, \infty\right)$

D: $I_\mu = \left(\overline{x} - \frac{s}{\sqrt{n}} \cdot \lambda_{\alpha}, \infty\right)$

E: $I_\mu = \left(-\infty,\ \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot \lambda_{\alpha}\right)$

F: $I_\mu = \left(-\infty,\ \overline{x} + \frac{s}{\sqrt{n}} \cdot \lambda_{\alpha}\right)$

Uppgift 10

Veckoutgifterna (i kr) för mat hos en viss familj kan uppfattas som observationer x_i av en $N(\mu, \sigma)$-fördelning där $\sigma = 300$ är känd. Utgifterna under en vecka antas oberoende av utgifterna under övriga veckor. Man vill utföra ett test på signifikansnivån 10% där man testar nollhypotesen $H_0 : \mu = 1800$ mot $H_1 : \mu < 1800$. Som testvariabel används ett medelvärde som är baserat på observationer av utgifterna under nio olika veckor. Stickprovsmedelvärdet beräknat på de nio observationerna blev $\bar{x} = 1700$.
Bestäm testets P-värde.

A: 0.841

B: 0.371

C: 0.317

D: 0.159

Uppgift 11

Man vill undersöka om följande data kommer från en binomialfördelning med parametern $n = 3$ och något värde på parametern p, dvs kan data komma från $Bin(3, p)$ för något p?

	0	1	2	3	Totalt
antal observationer	15	54	28	3	100

Om $p^*_{obs} = 0.3967$ används då man beräknar teststorheten Q fås att $Q = 6.5329$. Vilket av följande påståenden är sant?

A: H_0 kan varken förkastas på risknivån 1% eller risknivån 5%

B: H_0 kan både förkastas på risknivån 1% och risknivån 5%

C: H_0 kan förkastas på risknivån 1% , men inte på risknivån 5%

D: H_0 kan förkastas på risknivån 5% , men inte på risknivån 1%

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 ett 95% konfidensintervall för den effekt som belysningsindex har på jordgubbsvikten, dvs I_β, om man vet att effekten i fråga *inte* är signifikant på 5% nivå.

Ledning: Inga beräkningar behövs för att lösa uppgiften.

A: [6.81, 13.19]

B: [0.43, 3.57]

C: [−0.43, 4.43]

D: [−4.43, 0.43]

Del II

Uppgift 13

En spelare har två mynt i fickan, ett "riktigt" mynt med krona på ena sidan och klave på den andra och ett falskt mynt med krona på båda sidorna. Spelaren väljer slumpmässigt ett av mynten.

a) När myntet kastats visar det krona. Beräkna sannolikheten att det är det "riktiga" myntet. (5 p)

b) Antag att man kastar samma mynt en andra gång och får krona igen. Beräkna sannolikheten att det är det "riktiga" myntet. (5 p)

Uppgift 14

a) Ett mynt ger krona och klave med sannolikheterna p respektive $q = 1 - p$. Myntet kastas upprepade gånger tills man fått krona r gånger (man slutar alltså när krona kommer upp för r:te gången). Låt X beteckna antalet kast man gjort.

Bestäm väntevärde och varians för X uttryckta i parametrarna p och r. (5 p)

Ledning: Lättast är att utnyttja att man kan skriva X som $X = \sum_{i=1}^{r} U_i$, alltså som en summa av r stokastiska variabler $U_1, \ldots, U_r$ med samma kända fördelning (och vars väntevärde och varians alltså återfinns i formelsamlingen).

b) Förra julen gav ett föräldrapar sitt barn en basketkorg för inomhusbruk. Barnet blev mycket förtjust och övar skott flitigt. Så flitigt att föräldrarna nu förhandlat fram en begränsning: barnet får endast skjuta tills det satt 10 skott i korgen under en dag. Barnets träffsäkerhet har förbättrats under det gångna året och numera sitter vartannat skott i korgen.

Beräkna den approximativa sannolikheten att föräldrarna får höra bollen dunka i väggen (vilket den gör en gång för varje skott) fler än 7500 under ett år. Räkna på ett år som 365 dagar och antag att barnet bibehåller samma träffsäkerhet hela det kommande året. (5 p)

Ledning: Fördelningen från a-uppgiften är användbar. Om man inte lyckats lösa a-uppgiften kan man använda väntevärdet 21 och varians 13.6125 för antalet skott under en dag.

Uppgift 15

En forskare på en lantbrukshögskola vill undersöka om halten av quercetin, en antioxidant som anses motverka uppkomsten av bl. a. cancer, skiljer sig åt mellan två olika äppelsorter. Han mäter därför upp halten quercetin (i mg/100 g) i 8 äpplen av den första sorten $(x_1, \ldots, x_8)$ och 10 av den andra $(y_1, \ldots, y_{10})$, och räknar ut att $\overline{x} = \frac{1}{8}\sum_{i=1}^{8} x_i = 4.4723$, $\sum_{i=1}^{8}(x_i - \overline{x})^2 = 0.1277$, $\overline{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 4.6266$, och $\sum_{i=1}^{10}(y_i - \overline{y})^2 = 0.1997$. Mätvärdena kan antas vara observationer av oberoende stokastiska variabler $X_1, \ldots, X_8$ och $Y_1, \ldots, Y_{10}$. Vidare antas att $X_i \in N(m_1, \sigma)$, $i = 1, ..., 8$, samt att $Y_i \in N(m_2, \sigma)$, $i = 1, ..., 10$, där parametrarna m_1, m_2 och σ är okända.

a) Beräkna ett 95 % konfidensintervall för $m_1 - m_2$, och testa på signifikansnivån 5 % nollhypotesen $H_0 : m_1 = m_2$ mot alternativet $H_1 : m_1 \neq m_2$. Slutsatsen ska klart framgå. (4 p)

b) Antag nu, till skillnad från i (a)-delen, att det är känt från tidigare studier att $\sigma = 0.16$. Beräkna ett 95 % konfidensintervall för $m_1 - m_2$, och testa på signifikansnivån 5 % nollhypotesen $H_0 : m_1 = m_2$ mot alternativet $H_1 : m_1 \neq m_2$. Slutsatsen ska klart framgå. (3 p)

c) Ange den lägsta signifikansnivå på vilken vi kan förkasta nollhypotesen i b) givet de observationer vi har, dvs ange observerad signifikansnivå. Tänk på att testet är tvåsidigt! (3 p)

Uppgift 16

En cirkel med känd medelpunkt har den okända radien R_0. Inom cirkeln väljer man slumpmässigt n punkter (så varje punkt i cirkeln väljs med samma sannolikhet). Avstånden från punkterna till medelpunkten är $r_1, \ldots, r_n$.

a) Låt R vara den stokastiska variabel som beskriver avståndet från en i cirkel slumpmässigt vald punkt till cirkelns medelpunkt. Ange täthetsfunktionen för R. (2 p)

Ledning: För att se om man är på rätt väg kan man kontrollera att man får $E[R] = 2R_0/3$.

b) Visa att ML-skattningen av R_0 ges av

$$(R_0)^*_{obs} = \max_i \{r_i\}$$

Tydlig motivering krävs! (3 p)

c) Beräkna den korrigerade ML-skattningen av R_0, dvs en väntevärdesriktig skattning av R_0 på formen

$$C(n) \cdot (R_0)^*_{obs} = C(n) \max_i \{r_i\}$$

där $C(n)$ är en funktion av n. (5 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1917/SF1918/1919 SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 8 JANUARI 2020 KL 8.00–13.00.

Del I

Rätt rad:

1. C
2. 6
3. D
4. D
5. 0.26127
6. B
7. B
8. A
9. F
10. D
11. D
12. C

Kortfattade lösningar:

Uppgift 1

Vi använder unionsformeln:$P(A) + P(B) - P(A \cap B) = P(A \cup B)$

$P(A \cap B) = 1 - P(A^* \cup B^*) = 1 - 0.7 = 0.3.$

$P(A \cup B) = 1 - P(A^* \cap B^*) = 1 - 0.1 = 0.9$

Då får vi $P(A) + 0.6 - 0.3 = 0.9.$

Vilket ger $P(A) = 0.6$

Uppgift 2

Vi deriverar $F(x)$ för att få tätheten, $f(x) = F'(x) = 1/(2c)$ då $11-c \leq x \leq 11+c$. Därmed är X likformig fördelad U$(11-c, 11+c)$. Vi har, enligt formelsamling, $V(X) = (2c)^2/12 = 12$, dvs $c = 6$.

Uppgift 3

F.S. §2 ger $C(X,Y) = D(X)D(Y)\rho(X,Y) = \sqrt{2}\cdot\sqrt{5}\cdot(-\frac{2}{3})$.

$V(3X-4Y) = 3^2V(X) + (-4)^2V(Y) + 2C(3X,-4Y) = 9\cdot 2 + 16\cdot 5 - 2\cdot 3\cdot 4\cdot C(X,Y) =$

$= 18 + 80 - 24\sqrt{2}\cdot\sqrt{5}\cdot(-\frac{2}{3}) = 98 + 16\sqrt{10} = 148.6$

<u>Svar: 148.6</u>

Uppgift 4

Vi inför den stokastiska variabeln X, där X är hur många av de 10 följande kunderna som beställer kaffe.

Då gäller att $X \in Bin(10, 0.7)$ och vi söker $P(X \geq 5)$

För att kunna använda tabell 6 inför vi den stokastiska variabeln Y, där Y är hur många av de 10 följande kunderna som *inte* beställer kaffe.

Då gäller att $Y \in Bin(10, 0.3)$ och vi söker $P(Y \leq 5)$ Vilket enligt tabell 6 är 0.95265

<u>Alltså är rätt svar 0.95</u>

Uppgift 5

Eftersom summan av oberoende Poissonfördelade stokastiska variabler är Poissonfördelad inför vi den stokastiska variabeln $W = X + Y + Z$ där $W \in Po(1 + 0.5 + 0.9) = Po(2.4)$.

Vi söker nu $P(W = 2)$ Vi vill använda tabell 5 och använder oss då av att

$P(W = 2) = P(W \leq 2) - P(W \leq 1) = 0.56971 - 0.30884 = 0.26127.$

<u>Alltså är rätt svar 0.261</u>

Uppgift 6

Låt X vara en stokastisk variabel som beskriver variationen av poäng på antagningsprovet till utbildningen i fråga, $X \sim N(500, 100)$. Vi söker a sådant att $P(X \geq a) = 0.1$. Efter standardisering, har vi:

$$P\Big(Z \geq \frac{a-500}{100}\Big) = 0.1,$$

där $Z = (X - 500)/100 \sim N(0,1)$. Vi har nu, från Tabell 2, att $\frac{a-500}{100} = 1.2816$ och därmed $a = 500 + 1.2816 * 100 \approx 628$.

Uppgift 7

För att beräkna MK-skattningen av λ betraktar vi

$$Q(\lambda) = \sum_{i=1}^{3}\left(x_i - \frac{2}{\lambda}\right)^2.$$

Derivering map λ ger

$$\frac{dQ(\lambda)}{d\lambda} = \sum_{i=1}^{3} 2\left(x_i - \frac{2}{\lambda}\right)\frac{2}{\lambda^2}$$

Sätter vi derivatan till 0 fås ekvationen (använd $\lambda > 0$)

$$\sum_{i=1}^{3}\left(x_i - \frac{2}{\lambda}\right) = 0.$$

Löser man denna för λ får man att MK-skattningen av λ är

$$\lambda^*_{obs} = \underline{\frac{2}{\bar{x}} = 2.379.}$$

Uppgift 8

Detta är stickprov i par. Stickprovsmedelvärde och stickprovsstandardavviklesen för skillnader (före - efter) blir $\bar{z} = 11.6$ respektive $s_z = \sqrt{18.8}$. Eftersom populationernas standardavvikelser är okända, har vi följande eftersökta KI för genomsnittliga förändringen av kolesterolhalt:

$$\begin{aligned} I_{\mu_{fore}-\mu_{efter}}(0.95) =& \left(\bar{z} \pm t_{\alpha/2}(n-1)\frac{s_z}{\sqrt{n}}\right) \\ =& (11.6 - 2.78 * \sqrt{18.8}/\sqrt{5}; 11.6 + 2.78 * \sqrt{18.8}/\sqrt{5}) \\ =& (6.21, 16.99), \end{aligned}$$

$t_{\alpha/2}(n-1) = t_{0.025}(4) = 2.78$ är från Tabell 3, $\alpha = 0.05$, $n = 5$.

Uppgift 9

Vi vill förkasta nollhypotesen om små värden på $\overline{x}$ observeras, alltså måste vi använda ett uppåt begränsat intervall. Då σ är okänd måste den skattas med s. Alternativ F är således korrekt.

Uppgift 10

Vi har:

$$
\begin{aligned}
P\text{-varde} =& P\big(\bar{X} \leq \bar{x} | \mu = 1800\big) = |\bar{X} \sim N(\mu, \sigma/\sqrt{n})| \\
=& P\big(\bar{X} \leq 1700 | \bar{X} \sim N(1800, 300/\sqrt{9}\big) \\
=& \Phi((1700 - 1800)/100) = \Phi(-1) = 1 - \Phi(1) = |Tabell 1| \\
=& 1 - 0.8413 = 0.1587 \approx 0.159.
\end{aligned}
$$

Uppgift 11

Vi har här att $Q \sim \chi^2$-fördelad med $(4-1-1) = 2$ frihetsgrader eftersom vi har ett test av en fördelning med en skattad parameter. Således har vi att $Q \sim \chi^2(2)$ och eftersom det gäller att $\chi^2_{0.05}(2) = 5.99 < 6.5329 < 9.21 = \chi^2_{0.01}(2)$ kommer vi att förkasta H_0 på signifikansnivån 5%, men inte på signifikansnivån 1%.

Uppgift 12

Konfidensintervall I_β måste uppfylla:

- $\beta^*_{obs} = 2$, eftersom intervallet är symmetrisk runt β^*_{obs}, och
- noll, eftersom effekten som belysningsindex har på jordgubbsvikteninte inte är signifikant.

Det finns bara ett intervall (nämligen, i svarsalternativ C) som uppfyller de två kraven.

Del II

Uppgift 13

a) Metod 1.
Låt A och B vara "det visar krona" respektive "det är det "riktiga" myntet". Vi har $P(A) = 3/4$ (det är totalt fyra sidor på de två mynten, varav tre gynnar A), $P(B) = 1/2$ (det är två mynt). Vi har också att $P(A \cap B) = 1/4$ (det är bara en sida på de fyra totalt sidorna som gynnar både A och B). Detta ger

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{3/4} = 1/3.$$

Metod 2.
Vi begränsar hela utfallsrummet till "det visar krona", dvs till den händelse vi betingar på, och ansätter detta som ett nytt utfallsrum.

Det nya utfallsrummet består således av tre möjliga utfall (de tre sidorna med krona), varav bara ett utfall gynnar händelse "det är det "riktiga" myntet". Därmed är den sökta sannolikheten lika med 1/3.

b) Vi gör här på samma sätt som i a) Metod 2 ovan. Låt oss begränsa hela utfallsrummet till "det är en krona igen på andra kast av samma mynt", dvs till den händelse vi betingar på, och ansätter detta som ett nytt utfallsrum.

Det nya utfallsrummet består således av *fem möjliga utfall* (fyra utfall kan fås med det falska myntet som har krona på varje sida, nämligen {(Sid1, Sid1), (Sid1, Sid2), (Sid2, Sid1), (Sid2, Sid2)}) och bara ett utfall på det "riktiga" myntet som kan visa både krona och klave, nämligen (Kr, Kr)). Av de fem nämnda utfallen är det bara ett utfall som gynnar händelsen "det är det "riktiga" myntet". Därmed är den sökta sannolikheten lika med 1/5.

Uppgift 14

a) Vi har att

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

Där de stokastiska variablerna U_i, $i = 1, \ldots, r$ är oberoende och ffg-fördelade med parameter p. Enligt formelsamlingen gäller att

$$E(U_i) = \frac{1}{p} \qquad \text{och} \qquad V(U_i) = \frac{1-p}{(p)^2}.$$

Vi får därför att

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^{r} U_i\right) = \sum_{i=1}^{r} E(U_i) = r \cdot \frac{1}{p} = \frac{r}{p} \\ V(X) &= V\left(\sum_{i=1}^{r} U_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{r} V(U_i) = r \cdot \frac{1-p}{(p)^2} = \frac{r(1-p)}{(p)^2}. \end{aligned}$$

b) Låt X_i beteckna antal skott som behövs för att träffa korgen 10 gånger dag i. Då gäller att X_i ha den fördelning som behandlats i uppgift a, med $r = 10$ och $p = 1/2$. Totala antalet skott under ett år ges då av den stokastiska variabeln $Y = \sum_{i=1}^{365} X_i$. Då X_i:na är oberoende och likafördelade och summan innehåller många termer ger centrala gränsvärdessatsen att Y är approximativt normalfördelad med parametrar

$$\begin{aligned} E(Y) &= E\left(\sum_{i=1}^{365} X_i\right) = \sum_{i=1}^{365} E(X_i) = 365 \cdot \frac{10}{0.5} = 7300 \\ V(Y) &= V\left(\sum_{i=1}^{365} X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{365} V(X_i) = 365 \cdot \frac{10 \cdot 0.5}{(1-0.5)^2} = 7300 \\ D(Y) &= \sqrt{V(Y)} = \sqrt{7300} \end{aligned}$$

Sannolikheten att föräldrarna måste lyssna till fler än 7500 väggdunk under ett år blir därför

$$\begin{aligned} P(Y \geq 7500) &= 1 - P(Y \leq 7500) = 1 - P\left(\frac{Y - 7300}{\sqrt{7300}} \leq \frac{7500 - 7300}{\sqrt{7300}}\right) \\ &\approx 1 - \phi(2.34) \approx 1 - 0.9904 = 0.0096 \end{aligned}$$

Om man räknar med väntevärde 21 och varians 13.6125 Får man istället

$$\begin{aligned} P(Y \geq 7500) &= 1 - P(Y \leq 7500) = 1 - P\left(\frac{Y - 7665}{\sqrt{365 \cdot 13.6125}} \leq \frac{7500 - 7665}{\sqrt{4968.5625}}\right) \\ &\approx 1 - \phi(-2.34) \approx 1 - (1 - 0.9904) = 0.9904 \end{aligned}$$

Uppgift 15

a) Enligt F.S.§11.2 (d) gäller att

$$\frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{s\sqrt{\frac{1}{8} + \frac{1}{10}}} \sim t(8 + 10 - 2),$$

där

$$s^2 = \frac{(8-1)s_1^2 + (10-1)s_2^2}{8 + 10 - 2} = \frac{1}{16}\Big(\sum_{i=1}^{8}(X_i - \overline{X})^2 + \sum_{i=1}^{10}(Y_i - \overline{Y})^2\Big).$$

Kvantilernas definition ger

$$P\Big(-t_{0.025}(16) \leq \frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{s\sqrt{\frac{9}{40}}} \leq t_{0.025}(16)\Big) = 0.95,$$

och genom omskrivning av de båda olikheterna fås

$$P\Big(\overline{X} - \overline{Y} - t_{0.025}(16)s\sqrt{\frac{9}{40}} \leq m_1 - m_2 \leq \overline{X} - \overline{Y} + t_{0.025}(16)s\sqrt{\frac{9}{40}}\Big) = 0.95,$$

vilket ger konfidensintervallet

$$I_{m_1-m_2} = \overline{x} - \overline{y} \pm t_{0.025}(16)s\sqrt{\frac{9}{40}}$$

$$= -0.1543 \pm 2.12\sqrt{\frac{0.1277 + 0.1997}{16}}\sqrt{\frac{9}{40}} = \underline{-0.1543 \pm 0.1438}.$$

Eftersom $0 \notin I_{m_1-m_2}$, så förkastas H_0 (på nivån 5%).

b) Enligt F.S.§11.2 (a) gäller att

$$\frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{\sigma\sqrt{\frac{1}{8} + \frac{1}{10}}} \sim N(0,1),$$

där $\sigma = 0.16$. Kvantilernas definition ger

$$P\big(-\lambda_{0.025} \leq \frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{\sigma\sqrt{\frac{9}{40}}} \leq \lambda_{0.025}\big) = 0.95,$$

och genom omskrivning av de båda olikheterna fås

$$P\big(\overline{X} - \overline{Y} - \lambda_{0.025}\sigma\sqrt{\frac{9}{40}} \leq m_1 - m_2 \leq \overline{X} - \overline{Y} + \lambda_{0.025}\sigma\sqrt{\frac{9}{40}}\big) = 0.95,$$

vilket ger konfidensintervallet

$$I_{m_1-m_2} = \overline{x} - \overline{y} \pm \lambda_{0.025}\sigma\sqrt{\frac{9}{40}} = -0.1543 \pm 1.96 \cdot 0.16\sqrt{\frac{9}{40}} = \underline{-0.1543 \pm 0.1488}.$$

Eftersom $0 \notin I_{m_1-m_2}$, så förkastas H_0 (på nivån 5%).

c) Observerad signifikansnivå eller P-värdet är sannolikheten att få den observation vi fått eller något ännu mer extremt då H_0 sann. Vi söker därför $P(\bar{X} - \bar{Y} \leq \bar{x} - \bar{y})$ då $m_1 - m_2 = 0$. Vi gör om till $N(0,1)$ och får

$$P\left(\frac{\bar{X} - \bar{Y} - 0}{0.16 \cdot \sqrt{\frac{9}{40}}} \leq \frac{-0.1543 - 0}{0.16 \cdot \sqrt{\frac{9}{40}}}\right) = \Phi\left(-0.964375\sqrt{\frac{40}{9}}\right) = 1 - \Phi\left(0.964375\sqrt{\frac{40}{9}}\right) \approx$$
$$1 - \Phi(2.03) = 1 - 0.97882 = 0.02118$$

Eftersom vi har ett tvåsidigt test blir P-värdet dennna sannolikhets dubbla värde.

<u>Man kan förkasta på lägsta signifikansnivå 0.042</u>

Uppgift 16

a) Vi har att fördelningsfunktionen för R ges av

$$F_R(r) = P(R \leq r) = \frac{\pi r^2}{\pi R_0^2} = \frac{r^2}{R_0^2}, \qquad 0 \leq r \leq R_0$$

(se Definition 4.5 i kursboken). Genom att derivera får vi täthetsfunktionen

$$\underline{f_R(r)} = \frac{d}{dr}F_R(r) = \underline{\frac{2r}{R_0^2}, \qquad 0 \leq r \leq R_0.}$$

b) ML-skattningen maximerar likelihoodfunktionen

$$L(R_0) = f_{R_1}(r_1; R_0) f_{R_2}(r_2; R_0) \cdots f_{R_n}(r_n; R_0) = \prod_{i=1}^{n} f_{R_i}(r_i; R_0) = \prod_{i=1}^{n} \frac{2r_i}{R_0^2}.$$

Logaritmering ger

$$\ln L(R_0) = \ln \prod_{i=1}^{n} \frac{2r_i}{R_0^2} = \sum_{i=1}^{n} 2r_i - 2\ln R_0.$$

Vi ser nu att för att maximera $\ln L(R_0)$ så skall vi välja R_0 så litet som möjligt. Vi vet dock att R_0 måste vara större än den största observation vi fått, alltså får vi ML-skattningen

$$(R_0)^*_{obs} = \max_i \{r_i\}.$$

c) Vi behöver väntevärdet för den stokastiska variablen $Y = \max_i\{R_i\}$, där de stokastiska variablerna R_i är oberoende och har den täthetsfunktion som beräknats i a). Vi får att fördelningsfunktionen för Y ges av

$$\begin{aligned} F_Y(t) &= P(Y \leq t) = P(\max_{1\leq i \leq n}\{R_i\} \leq t) = \\ &= P(R_1 \leq t, R_2 \leq t, \ldots, R_n \leq t) = \{\text{oberoende}\} = \textstyle\prod_{i=1}^{n} P(R_i \leq t) = (F_R(t))^n \end{aligned}$$

för $0 \leq r \leq R_0$. Täthetsfunktionen för Y blir därför

$$f_Y(t) = \frac{dF_Y(t)}{dt} = n(F_R(t))^{n-1}\frac{dF_R(t)}{dt} = n(F_R(t))^{n-1} f_R(t)$$

för $0 \leq r \leq R_0$. Väntevärdet för Y fås nu som

$$\begin{aligned} E(Y) &= \int_0^{R_0} y f_Y(y)dy = \int_0^{R_0} yn(F_R(y))^{n-1} f_R(y)dy = \int_0^{R_0} yn(\frac{y^2}{R_0^2})^{n-1}\frac{2y}{R_0^2}dy \\ &= \int_0^{R_0} 2n\frac{y^{2n}}{R_0^{2n}}dy = \frac{2n}{R_0^{2n}}\left[\frac{y^{2n+1}}{2n+1}\right]_0^{R_0} = \frac{2n}{2n+1}R_0. \end{aligned}$$

Således ges den korrigerade ML-skattningen av

$$\widehat{R}_0 = \underline{\frac{2n+1}{2n} \cdot \max_i\{r_i\}.}$$