Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1923/SF1924 SANNOLIKHETSTEORI OCH STATISTIK,
MÅNDAG 15 AUG 2022 KL 8.00–13.00.

Examinator: Björn-Olof Skytt, 08-790 86 49.

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 får även tillgodoräkna sig uppgift 12 och behöver alltså ej besvara den. Gränsen för godkänt är 9 poäng. Möjlighet att komplettera ges för tentander med 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 del I och poäng på del II krävs för högre betyg än E. På denna del skall resonemang och uträkningar 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 4 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

I en gjord undersökning om dödsstraff i U.S.A. svarade 64% att de stödjer dödsstraff. Kvinnorna utgjorde 48% av de tillfrågade. Av kvinnorna svarade 46% att de var för dödsstraff. Hur stor är sannolikheten att en slumpmässigt vald person är en man som vill ha dödsstraff?

A: 0.2800

B: 0.4192

C: 0.6550

D: 0.8062

Uppgift 2

$E(X) = 3$, $E(Y) = 2$, $D(X) = 2$, $D(Y) = 1$, $C(X, Y) = 1$.
Beräkna $E[(2X - 3Y)^2]$

A: 7

B: 13

C: 19

D: 25

Uppgift 3

De stokastiska variablerna X och Y har den simultana sannolikhetsfunktionen

$p_{X,Y}(x, y)$	$X = 1$	$X = 2$
$Y = 2$	0.2	0.5
$Y = 4$	0.2	0.1

Beräkna $V(2X - Y)$.

Uppgift 4

En kontinuerlig stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ 1 - e^{-0.3x} & \text{om } x \geq 0. \end{cases}$$

Beräkna $P(X > 5 | X > 2)$.

Uppgift 5

Låt X och Y vara två oberoende stokastiska variabler sådana att $X \in N(2.8, 0.5)$ och $Y \in N(2.8, 0.5)$. Z=max(X,Y). Beräkna $P(Z > 3)$.

A: 0.43

B: 0.48

C: 0.52

D: 0.57

Uppgift 6

Antag att det till en vårdcentral kommer i genomsnitt en patient om dagen som skall provtas för streptokockinfektion. Antag vidare att antalet patienter som behöver provtas för streptokocker en viss dag är poissonfördelat och att antalet patienter en dag är oberoende av antalet patienter andra dagar. Vårdcentralen hade i början av året 240 testkit för streptokockinfektion. Beräkna approximativt sannolikheten att testkiten kommer att räcka hela året om ett år anses bestå av 256 arbetsdagar.

A: 0.16

B: 0.48

C: 0.52

D: 0.84

Uppgift 7

X, Y, Z är oberoende stokastiska variabler. $X \in Exp(\lambda)$, $Y \in Exp(3\lambda)$, $Z \in Exp(4\lambda)$. Vi har fått utfallen $x = 8$, $y = 3$, $z = 2$. Bestäm ML-skattningen av λ.

A: 0.067

B: 0.094

C: 0.12

D: 0.23

Uppgift 8

Låt $\bar{x} = 137.0$, $\bar{y} = 208.0$, $\sigma_x^2 = 92.5$, $,\sigma_y^2 = 163.0$, $n_x = 5$ och $n_y = 5$ vara givet och antag att $X_i \in N(\mu_x, \sigma_x)$ och $Y_i \in N(\mu_y, \sigma_y)$ samt att alla dessa stokastiska variabler är oberoende. Ange övre gränsen för det 95%-iga ensidigt uppåt begränsade konfidensintervallet $I_{\mu_x - \mu_y}$.

A: -50.2

B: -57.0

C: -57.7

D: -59.2

Uppgift 9

Låt datamängden 4, 9, 9, 13, 6 vara given. Varje data anses vara ett utfall av en stokastisk variabel X_i, där X_i:na antas vara oberoende och Poissonfördelade med väntevärde μ. D.v.s. varje $X_i \in Po(\mu)$. Eftersom X_i:na är Poissonfördelade skattar vi $E(X_i) = \mu$ med $\bar{x}$ och $D(X_i) = \sqrt{\mu}$ med $\sqrt{\bar{x}}$.
Ange medelfelet för skattningen av μ.

A: 1.28

B: 1.43

C: 2.81

D: 3.67

Uppgift 10

Antag att $X \in \text{ffg}(p)$ och låt H_0 vara att $p = 0.1$. Vi vill testa H_0 mot alternativet $H_1 : p = 0.4$ och förkastar H_0 till förmån för H_1 om vi observerar ett litet värde x på X.
Antag att vi observerat $x = 3$. Bestäm testets P-värde.

A: 0.180

B: 0.271

C: 0.640

D: 0.784

Uppgift 11

Produktionsuppgifter visar att i normal drift för ett visst märke av datorhårddisk har 95 % inga fel, 3 % har ett fel och 2 % har mer än ett fel. För ett slumpmässigt urval av 400 av dessa hårddiskar från en veckas produktion visade sig 375 ha inga fel, 15 ha ett fel och 10 ha mer än ett fel.Testa på 5%-nivån nollhypotesen H_0 att kvaliteten på utdata från denna vecka överensstämmer med det vanliga mönstret.

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

Ett sätt att presentera mätdata är att göra en så kallad boxplott. Andelen av de uppmätta data som då ligger i själva boxen skall då vara:

A: 50%

B: 90%

C: 95%

D: 99%

Del II

Uppgift 13

För att undersöka sambandet mellan barns födelsevikt och längden på motsvarande graviditet har man registrerat värden vid 18 födslar. Låt $x_1, \dots, x_{18}$ vara längden (i dagar) för de graviditeter som varit en del av studien och låt $y_1, \dots, y_{18}$ vara motsvarande födelsevikter (kg).
Vi önskar nu göra en enkel linjär regression enligt formen $Y_i = \alpha + \beta x_i + \varepsilon_i$, där ε_i som vanligt är oberoende $N(0, \sigma)$-fördelade variabler. För att kunna bilda en sådan modell har man tagit fram följande värden från observationerna:

$$\sum_i x_i = 4902, \quad \sum_i y_i = 60.35, \quad \sum_i x_i^2 = 1342700, \quad \sum_i y_i^2 = 209.05, \quad \sum_i x_i y_i = 16630.$$

a) Tag fram punktskattningar för α och β i regressionslinjen. (4 p)

b) Undersök på nivån 5% om lutningen β bör vara med i modellen. (4 p)

c) Bestäm ett konfidensintervall med konfidensgrad 99% för den genomsnittliga födelsevikten. (2 p)

Uppgift 14

För brandstationer i två distrikt är antalet bränder som man behöver rycka ut till under en vecka Poissonfördelade med parametrarna $\mu_1 = 3$ ("Station 1") respektive $\mu_2 = 5$ ("Station 2").

a) Vad är sannolikheten att de bägge stationerna kombinerat har 3 eller färre utryckningar under en vecka? (3 p)

b) Givet att det kombinerade antalet utryckningar under en vecka var 6, vad är sannolikheten att man från Station 1 hade fler än 3 utryckningar? (7 p)

Uppgift 15

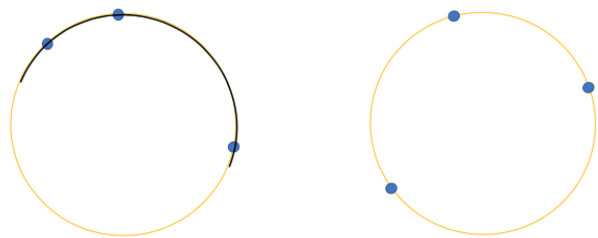
Vi har ett stickprov $X_1, \dots, X_n$ från $N(\mu, \sigma)$ där både μ och σ är okända.

a) Definiera en lämplig punktskattning av σ^2 och konstruera ett nedåt begränsat konfidensintervall, konfidensgrad 95%, för σ^2. Svaret får uttryckas i den skattning du definierat. (3 p)

b) Du vill nu testa nollhypotesen $H_0 : \sigma^2 = 1$ mot $H_1 : \sigma^2 > 1$. Om testet ska utföras på nivån 5%, hur stor måste stickprovsstorleken n vara för att testets styrka ska vara minst 95% om det sanna värdet är $\sigma^2 = 4$? (7 p)

Uppgift 16

a) På en cirkel placerar vi ut tre punkter på måfå, det vill säga likformigt över cirkelns omkrets. Beräkna sannolikheten att punkterna placeras på ett sätt så att en halvcirkeln kan täcka dem; se Figur 1 för en illustration. (5 p)



Figur 1: I vänstra figuren kan de tre punkterna täckas av en halvcirkel, i högra figuren är det ej möjligt.

b) Betrakta nu kvadraten $D = [-1, 1] \times [-1, 1]$. Två punkter placeras ut slumpmässigt inuti D. Beräkna sannolikheten att bägge punkterna kan täckas (samtidigt) av en kvadrat med sidan 1. (5 p)

Lagen om total förväntan kan vara användbar: För en händelse H och en kontinuerlig stokastisk variabel X, med möjliga utfall beskrivna av mängden $\mathcal{X}$ och täthetsfunktion f_X, säger den att

$$P(H) = \int_{\mathcal{X}} P(H|X = x) f_X(x) dx.$$

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1923/1924 SANNOLIKHETSTEORI OCH STATISTIK, MÅNDAG 15 AUGUSTI 2022 KL 8.00–13.00.

Del I

Del I, Svar:

1. B

2. B

3. 2.44

4. 0.407

5. D

6. A

7. C

8. D

9. A

10. B

11. A

12. A

Del I, Lösningsförslag:

Uppgift 1

åt K vara händelsen att en kvinna är tillfrågad, och låt M vara händelsen att det är man som är tillfrågad. Låt D vara händelsen att en tillfrågad person är för dödsstraff. Vi har då, enligt problemlydelsen att $P(D|K) = 0.46$, $P(K) = 0.48$, och $P(D) = 0.64$.
Vi söker händelsen $P(M \cap D)$.
$P(D) = P(M \cap D) + P(K \cap D) = P(M \cap D) + P(D|K)P(K)$
$\Rightarrow P(M \cap D) = 0.64 - 0.46 \cdot 0.48 = 0.4192$

Uppgift 2

Givet: $E(X) = 3$, $E(Y) = 2$, $D(X) = 2$, $D(Y) = 1$, $C(X,Y) = 1$.
$E[(2X - 3Y)^2] = E[4X^2] - E[12XY] + E[9Y^2] =$
$= 4(E^2[X] + V[X]) - 12(C[X,Y] + E[X]E[Y]) + 9((E^2[Y] + V[Y]) =$
$= 4(9 + 4) - 12(1 + 3 \cdot 2) + 9(4 + 1) = 52 - 84 + 45 = 13$

Uppgift 3

Sätt $2X - Y = Z$. Då fås $p_Z(-2) = 0.2, p_Z(0) = 0.3, p_Z(2) = 0.5$.
$V(Z) = E(Z^2) - E^2(Z)$

$$E(Z) = \sum_{i=1}^{n} z \cdot p_Z(z) = -2 \cdot 0.2 + 0 \cdot 0.3 + 2 \cdot 0.5 = 0.6$$

$$E(Z^2) = \sum_{i=1}^{n} z^2 \cdot p_Z(z) = (-2)^2 \cdot 0.2 + 0^2 \cdot 0.3 + 2^2 \cdot 0.5 = 2.8$$

$\Rightarrow V(Z) = 2.8 - 0.6^2 = 2.44$

Uppgift 4

$$P(X > 5|X > 2) = \frac{P(X > 5 \cap X > 2)}{P(X > 2)} = \frac{P(X > 5)}{P(X > 2)} = \frac{1 - F_X(5)}{1 - F_X(2)} = \frac{e^{-0.3 \cdot 5}}{e^{-0.3 \cdot 2}} = e^{-0.9} = 0.407$$

Uppgift 5

$P(max(X,Y) > 3) = 1 - P(X < 3)PY < 3)$

$$P(X < 3) = PY < 3) = P(\frac{X - 2.8}{0.5} < \frac{3 - 2.8}{0.5}) = \Phi(0.4) = 0.6554$$

$\Rightarrow P(max(X,Y) > 3) = 1 - 0.6554^2 = 0.57$

Uppgift 6

Låt X_i vara antal anländande patienter dag i. $X_i \in Po(1)$.Låt Y vara antal anländande patienter under arbetsåret. Eftersom summan av oberoende Poissonfördelade stokastiska variabler är Poissonfördelad gäller att

$$Y = \sum_{i=1}^{256} X_i \in Po(256)$$

Nu gäller enligt §6 i F.S. att Y är approximativt Normalfördelad eftersom $256 > 15$. Dvs. $Y \sim N(256, \sqrt{256})$.
Om testkiten ska räcka måste gälla att $Y \leq 240$

$$P(Y \leq 240 = P(\frac{Y - 256}{\sqrt{256}} < \frac{240 - 256}{\sqrt{256}}) \approx \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.16$$

Uppgift 7

Notationen $X \in \mathrm{Exp}(\lambda)$,$Y \in \mathrm{Exp}(3\lambda)$,$Z \in \mathrm{Exp}(4\lambda)$ betyder att tätheten för X är $f_X(x) = \lambda e^{-\lambda x}$,för Y är $f_Y(y) = 3\lambda e^{-3\lambda y}$,för Z är $f_Z(z) = 4\lambda e^{-4\lambda z}$. Därmed blir likelihoodfunktionen

$$L(\lambda) = \lambda e^{-\lambda x} \cdot 3\lambda e^{-3\lambda y} \cdot 4\lambda e^{-4\lambda z} = 12\lambda^3 e^{-\lambda(8+3\cdot3+4\cdot2)}$$

Då blir log-likelihoodfunktionen

$$\ln L(\lambda) = \ln 12 + 3\ln\lambda - 25\lambda.$$

Om vi maximerar $\ln L(\lambda)$ m a p λ har vi

$$\frac{d\ln L(\lambda)}{d\lambda} = \frac{3}{\lambda} - 25 = 0 \Leftrightarrow \lambda = \frac{3}{25} = 0.12.$$

ML-skattningen $= 0.12$

Uppgift 8

Vi ska bilda ett konfidensintervall för skillnaden mellan två Normalfördelade stickprovs väntevärden då standardavvikelserna är kända. Vi kan då ta hjälp av §12.1 och §11.3 i F.S och får då det tvåsidiga konfidensintervallet.

$$I_{\mu_x-\mu_y} = \bar{x} - \bar{y} \pm \lambda_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{5} + \frac{\sigma_y^2}{5}}$$

Men nu ska vi ha ett ensidigt uppåt begränsat konfidensintervall och då får vi

$$I_{\mu_x-\mu_y} = \left(-\infty, \bar{x} - \bar{y} + \lambda_\alpha \sqrt{\frac{\sigma_x^2}{5} + \frac{\sigma_y^2}{5}}\right)$$

Eftersom vi ska ha konfidensgrad 95% fås $\lambda_\alpha = \lambda_{0.05} = 1.6449$

$$\Rightarrow I_{\mu_x-\mu_y} = \left(-\infty, 137 - 208 + 1.6449\sqrt{\frac{92.5}{5} + \frac{163.0}{5}}\right) = (-\infty, -59.2)$$

Uppgift 9

Enligt §9.3 i F.S så definieras medelfelet för skattningen av θ som skattningen av $D(\theta^*)$. I vårt fall är $\theta = \mu$.
Eftersom

$$D(\mu^*) = D(\bar{X}) = \frac{D(X_i)}{\sqrt{n}} = \frac{\sqrt{\mu}}{\sqrt{n}}$$

så blir skattningen av $D(\mu^*)$ lika med

$$D^*_{obs}(\mu^*) = \left(\frac{\sqrt{\mu}}{\sqrt{n}}\right)^*_{obs} = \frac{\sqrt{\mu^*_{obs}}}{\sqrt{n}} = \frac{\sqrt{\bar{x}}}{\sqrt{n}} = \frac{\sqrt{8.2}}{\sqrt{5}} = 1.28$$

Uppgift 10

P-värdet är P(förkasta H_0) om H_0 är sann utgående från de observationer eller den obsevation man fått. I vårt fall blir detta $P(X \leq 3)$ om $X \in ffg(0.1)$
Då blir P-värdet $p_X(1) + p_X(2) + p_X(3) = 0.1 + 0.9 \cdot 0.1 + 0.9^2 \cdot 0.1 = 0.271$

Uppgift 11

Här har vi test av given fördelning. Se § 14.3
Om $np_i \geq 5$ för alla i gäller att $Q \sim \chi^2$-fördelad, där

$$Q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$$

Här gäller att $np_i \geq 5$ för alla i.

$$Q = \frac{(375 - 400 \cdot 0.95)^2}{400 \cdot 0.95} + \frac{(15 - 400 \cdot 0.03)^2}{400 \cdot 0.03} + \frac{(10 - 400 \cdot 0.02)^2}{400 \cdot 0.02} = 1.3158$$

Vi ska jämföra Q med $\chi^2_\alpha(r-1)$ där r = 3 i vårt fall och $\alpha = 0.05$ respektive 0.01.
$\chi^2_{0.05}(2) = 5.99$ och $\chi^2_{0.01}(2) = 9.21$
Om $Q < \chi^2_\alpha(r-1)$ så förkastas inte H_0. Alltså kan H_0 varken förkastas på risknivån 1% eller risknivån 5% i detta fall.

Uppgift 12

Rätt svar är 50%. Se tex. läroboken Figur 10.3 på sidan 231.

Del II

Uppgift 13

a) För att hitta punktskattningar $\hat{\alpha}, \hat{\beta}$ av koefficienterna i regressionslinjen använder vi MK-skattningarna:

$$\begin{aligned}\hat{\beta} &= \frac{S_{xy}}{S_{xx}},\\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}.\end{aligned}$$

Med givna värden får vi för $\hat{\beta}$

$$\begin{aligned}\hat{\beta} &= \frac{S_{xy}}{S_{xx}}\\ &= \frac{\sum x_k y_k - \frac{1}{n}\sum x_k \sum y_k}{\sum x_k^2 - \frac{1}{n}\left(\sum x_k\right)^2}\\ &= \frac{16630 - \frac{1}{18}4902 \cdot 60.35}{1342700 - \frac{1}{18}4902^2}\\ &= 0.0252.\end{aligned}$$

Vidare fås då för $\hat{\alpha}$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{1}{18}\left(60.35 - 0.0252 \cdot 4902\right) = -3.51.$$

<u>Svar:</u> Punktskattningarna ges av $\hat{\alpha} = -3.51$ och $\hat{\beta} = 0.0252$.

b) Vi använder oss här av konfidensmetoden för att testa

$$H_0: \ \beta = 0,$$

vilket svarar mot att x-termen ej ska vara med i modellen (lutning 0), mot

$$H_1: \ \beta \neq 0.$$

Ett konfidensintervall med konfidensgrad $100(1-p)\%$ för β ges av (se FS avsnitt 13), med n observationer,

$$\hat{\beta} \pm t_{p/2}(n-2)\frac{s}{\sqrt{S_{xx}}},$$

där s kan beräknas som, t.ex.,

$$(n-2)s^2 = S_{yy} - \hat{\beta}^2 S_{xx},$$

där i sin tur högerledet kan fås frå de givna värdena (see FS13.3).

Med insättning av värdena för $\sum x_k$, $\sum x_k^2$, $\sum y_k$ och $\sum y_k^2$ har vi

$$s^2 = 0.0968,$$

Konfidensintervallets undre och övre gräns ges därför av

$$\begin{aligned}\hat{\beta} \pm t_{p/2}(n-2)\frac{s}{\sqrt{S_{xx}}} &= 0.0252 \pm t_{0.025}(16) \cdot \frac{\sqrt{0.0968}}{\sqrt{7722}} \\ &= 0.0252 \pm 2.12 \cdot 0.0035 \\ &= 0.0252 \pm 0.0078,\end{aligned}$$

vilket ger $I_\beta = [0.0174, 0.033]$. Eftersom $0 \notin I_\beta$ kan vi på nivån 5% förkasta H_0 mot den alternativa hypotesen $\beta \neq 0$.

Svar: På nivå 5% bör lutningen vara med i modellen.

c) Givet vår modell så ges ett konfidensintervall, med konfidensgrad $100(1-p)\%$, för den genomsnittliga födelsevikten av

$$\hat{\alpha} + \hat{\beta}\bar{x} \pm t_{p/2}(n-2)\frac{s}{\sqrt{n}},$$

vilket här svarar mot, med $p = 0.01$,

$$\begin{aligned}&-3.51 + 0.0252 \cdot \frac{4902}{18} \pm 2.92 \cdot \frac{\sqrt{0.0968}}{\sqrt{18}} \\ &= 3.35 \pm 2.92 \cdot 0.073 \\ &= 3.35 \pm 0.21.\end{aligned}$$

Konfidensintervallet ges alltså av $[3.14, 3.56]$.

Svar: Ett konfidensintervall med konfidensgrad 99% ges av $[3.14, 3.56]$.

Uppgift 14

Låt X_1 och X_2 vara antalet bränder i respektive distrikt under en vecka. Vi har att $X_1 \in Po(\mu_1)$, $X_2 \in Po(\mu_2)$.

a) Låt Y beteckna det totala antalet bränder i de två distrikten under en vecka: $Y = X_1 + X_2$. Då X_1 och X_2 kan betraktas som oberoende stokastiska variabler har vi, från Poissonfördelningens egenskaper,

$$Y = X_1 + X_2 \in Po(\mu_1 + \mu_2).$$

Därmed gäller att

$$\begin{aligned}P(Y \leq 3) &= \sum_{k=0}^{3} \frac{(\mu_1 + \mu_2)^k e^{-\mu_1 - \mu_2}}{k!} \\ &= \sum_{k=0}^{3} \frac{8^k e^{-8}}{k!} \\ &= 0.042.\end{aligned}$$

Svar: Sannolikheten att de bägge stationerna kombinerat har 3 eller färre utryckningar är 0.042.

b) Vi söker nu den betingade sannolikheten $P(X_1 > 3|Y = 6)$. Vi ger här två lösningar, en baserad på definitionen av betingad sannolikhet och sannolikhetsfunktionen för Poissonfördelningen. Den andra på en specifik egenskap hos just Poissonfördelningen.

Alternativ 1: Låt Z_1, Z_2 vara två oberoende Poisson-fördelade variabler med parameter λ_1 respektive λ_2. För några värden $n, m \in \mathbb{N}$, $n \geq m$, antag att vi söker $P(Z_1 > m|Z_1+Z_n = n)$; notera att uppgiften är ett speciallfall av denna formuleringen. Då Z_1, Z_2 är oberoende har vi $Z_1 + Z_2 \in Po(\lambda_1 + \lambda_2)$.

Låt oss nu betrakta den betingade fördelningen för Z_1, betingat på $Z_1 + Z_2$.

$$\begin{aligned}
P(Z_1 = m|Z_1 + Z_2 = n) &= \frac{P(Z_1 = m \cap Z_1 + Z_2 = n)}{P(Z_1 + Z_2 = n)} \\
&= \frac{P(Z_1 = m \cap Z_2 = n - m)}{P(Z_1 + Z_2 = n)} \\
&= \frac{\frac{\lambda_1^m e^{-\lambda 1} \lambda_2^{n-m} e^{-\lambda 2}}{m!(n-m)!}}{\frac{(\lambda_1+\lambda_2)^n e^{-\lambda_1-\lambda_2}}{n!}} \\
&= \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^m \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-m} \frac{n!}{m!(n-m)!}.
\end{aligned}$$

Definiera nu $p = \lambda_1/(\lambda_1 + \lambda_2)$. Då har vi att $\lambda_2/(\lambda_1 + \lambda_2) = 1 - p$ och uttrycket ovan kan skrivas

$$p^m(1-p)^{n-m}\binom{n}{m}.$$

Det är inget annat än sannolikhetsfunktionen för binomialfördelningen med parametrar $p = \lambda_1/(\lambda_1 + \lambda_2)$ och n. Det vill säga den betingade fördelningen för Z_1 givet $Z_1 + Z_2$ är

$$Bin(\lambda_1/(\lambda_1 + \lambda_2), Z_1 + Z_2).$$

Översatt till vårt specifika problem så har vi $\lambda_1 = \mu_1 = 3$, $\lambda_2 = \mu_2 = 5$, vilket ger $p = 3/8$ $n = 6$ och vi är intresserade av $m = 4, 5, 6$ (eftersom > 3). Tag $X \in Bin(3/8, 6)$. Då söker vi

$$P(X > 3) = 0.15.$$

Alternativ 2 Per definition ges den eftersökta betingade sannolikheten av

$$\frac{P(X_1 > 3 \cap \; X_1 + X_2 = 6)}{P(X_1 + X_2 = 6)}.$$

Sannolikheten i täljaren ges av

$$\begin{aligned}
P(X_1 > 3 \cap X_1 + X_2 = 6) &= P(\{X_1 = 4 \cap X_2 = 2\} \cup \{X_1 = 5 \cap X_2 = 1\} \cup \{X_1 = 6 \cap X_2 = 0\}) \\
&= P(X_1 = 4 \cap X_2 = 2) + P(X_1 = 5 \cap X_2 = 1) + P(X_1 = 6 \cap X_2 = 0) \\
&= P(X_1 = 4)P(X_2 = 2) + P(X_1 = 5)P(X_2 = 1) + P(X_1 = 6)P(X_2 = 0) \\
&= 0.018.
\end{aligned}$$

där vi i andra steget använt definitionen av betingad sannolikhet och i det tredje att X_1 och X_2 är oberoende. På samma sätt har vi för termen i nämnaren:

$$P(X_1 + X_2 = 6) = \sum_{i=0}^{6} P(X_1 = i)P(X_2 = 6 - i) = 0.12.$$

Därmed får vi

$$P(X_1 > 3|Y = 6) = \frac{0.018}{0.12} = 0.15.$$

Svar: Sannolikheten att man från Station 1 hade fler än 3 utryckningar, givet totalt 6st, är 0.15.

Uppgift 15

a) Som stickprovsvariabel kan vi välja stickprovsvariansen: $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$, där $\bar{X}$ är det aritmetiska medelvärdet av observationerna $X_1, \ldots, X_n$. Då vi har ett oberoende, normalfördelat stickprov gäller enligt §11.1b) att

$$\frac{(n-1)S^2}{\sigma^2} \in \chi^2(n-1).$$

Alternativ 1: För att skapa det efterfrågade ensidigt nedåt begränsade konfidensintervallet, tag hjälp av § 12.4 och §11.b och byt $\frac{\alpha}{2}$ mot α i nedre gränsen och sätt övre gränsen till ∞. Då fås med s^2 punktskattningen som svarar mot stickprovsvariabeln S^2 har vi alltså konfidensintervallet

$$\left(\frac{(n-1)s^2}{\chi^2_{0.05}(n-1)}, \infty\right).$$

Alternativ 2: Eftersom

$$P(Z > \chi^2_\alpha(n-1)) = \alpha = 0.05$$

där $Z \in \chi^2(n-1)$ gäller att

$$\begin{aligned} 0.95 &= P\left(\frac{(n-1)}{\sigma^2}S^2 \leq \chi^2_{0.05}(n-1)\right) \\ &= P\left(\sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{0.05}(n-1)}\right). \end{aligned}$$

Med s^2 punktskattningen som svarar mot stickprovsvariabeln S^2 har vi alltså konfidensintervallet

$$\left(\frac{(n-1)s^2}{\chi^2_{0.05}(n-1)}, \infty\right).$$

Svar: Konfidensintervallet ges av $\left(\frac{(n-1)s^2}{\chi^2_{0.05}(n-1)}, \infty\right)$.

b) Styrkan hos testet är P(förkasta H_0) då H_1 är sann.

Vi skattar σ^2 med s^2 och förkastar $H_0 : \sigma^2 = 1$ till förmån för $H_1 : \sigma^2 = 4$ om s^2 är tillräckligt mycket större än 1. Det är det enligt kriteriet i a-uppgiften om

$$s^2 > \frac{(n-1)s^2}{\chi^2_\alpha(n-1)}$$

Givet konfidensintervallet i (a) vill vi alltså att följande ska gälla:

$$P\left(\frac{(n-1)S^2}{\chi^2_\alpha(n-1)} > 1|\sigma^2 = 4\right) \geq 0.95,$$

Eftersom

$$\frac{(n-1)S^2}{\sigma^2} \in \chi^2(n-1)$$

gäller att

$$P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{\chi^2_\alpha(n-1)}{\sigma^2}|\sigma^2 = 4\right) = P\left(\frac{(n-1)S^2}{4} > \frac{\chi^2_{0.05}(n-1)}{4}\right) \geq 0.95$$

Dvs. det ska gälla att

$$\frac{\chi^2_{0.05}(n-1)}{4} < \chi^2_{0.95}(n-1)$$

I χ^2-tabellen ser vi att det krävs att $n = 13$, eller större, för att ovan olikhet ska gälla.

<u>Svar:</u> Stickprovsstorleken behöver vara $n = 13$ eller större.

Uppgift 16

a) Vi kan anta att punkterna placeras ut på enhetscirkeln och deras placering anges som avståndet medurs från x-axeln. Laat $X, Y, Z \in [0, 2\pi)$ beteckna positionerna för de tre punkterna. Eftersom att rotera cirkeln inte påverkar punkternas inbördes placring eller avsånd kan vi dessutom anta att en av punkterna, säg X, ligger på x-axeln, så att $X = 0$.

Vi kan utnyttja symmetrin ytterligare genom att anta att $Y \in [0, \pi]$—skulle Y-punkten vara på ett avstånd längre bort från X kan vi istället betrakta spegelbilden, med avseende på x-axeln. Vi har därmed

$$f_Z(z) = \frac{1}{2\pi}, \quad z \in [0, 2\pi],$$
$$f_Y(y) = \frac{1}{\pi}, \quad y \in [0, \pi].$$

Låt H beteckna händelsen av intresse: att samtliga tre punkter kan täckas av en halvcirkel. För att beräkna motsvarande sannolikhet börjar vi med att betinga på Ys placering: Givet att $Y = y$, för något $y \in [0, \pi]$, gäller att H inträffar om Z placeras inom $[0, \pi]$ (inom halvcrikeln medurs från x-axeln) eller inom $[\pi + y, 2\pi]$. Det senare ges av att vi skapar en halvcirkeln som startar som mest $\pi - y$ moturs från x-axeln och sen täcker y—punkten ska alltså ligga inom $[2\pi - (\pi - y), 2\pi] = [\pi + y, 2\pi]$. Vi har därför den betingade sannolikheten

$$\begin{aligned} P(H|Y = y) &= P\left(\{Z \in [\pi + y, 2\pi]\} \cup \{Z \in [0, \pi]\}\right) \\ &= \frac{2\pi - (\pi + y)}{2\pi} + \frac{\pi}{2\pi} \\ &= \frac{2\pi - y}{2\pi}. \end{aligned}$$

För att beräkna den obetingade sannolikheten använder vi oss nu av lagen om total förväntan:

$$\begin{aligned} P(H) &= \int_0^\pi P(H|Y = y) f_Y(y) dy \\ &= \int_0^\pi \frac{2\pi - y}{2\pi} \times \frac{1}{\pi} dy \\ &= \frac{3}{4}. \end{aligned}$$

<u>Svar:</u> Sannolikheten att alla tre punkter kan täckas med en halvcirkel är 3/4.

b) Vi betecknar de två punkterna som placeras ut med $X = (X_1, X_2)$ och $Y = (Y_1, Y_2)$. Då gäller enligt lydelsen att X_1, X_2, Y_1, Y_2 är oberoende stokastiska variabler och alla likformigt fördelade på $[-1, 1]$.

Liknande del (a) kan vi utnyttja problemets symmetri: vi kan anta att X_1 och X_2 är likformigt fördelade på $[0, 1]$ (spegla annars punktens placering till första kvadranten). Låt återigen H beteckna händelsen av intresse. Vi använder nu samma strategi som i (a): betinga på ett specifikt utfall för X och använd sedan lagen om total förväntan.

Då vi nu antar att X ligger i första kvadranten, för en given punkt $(X_1, X_2) = (x_1, x_2)$ gäller att H inträffar om $Y_1 > x_1 - 1$ och $Y_2 > x_2 - 1$. Det ger den betingade sannolikheten

$$\begin{aligned} P(H|X = (x_1, x_2)) &= P(Y_1 > x_1 - 1, Y_2 > x_2 - 1 | X = (x_1, x_2)) \\ &= \int_{x_1 - 1}^1 \frac{1}{2} dy_1 \int_{x_2 - 1}^1 \frac{1}{2} dy_2 \\ &= \frac{(2 - x_1)(2 - x_2)}{4}, \end{aligned}$$

där vi använt att Y_1 och Y_2 är oberoende och likformigt fördelade på $[-1, 1]$, så att de har täthet $f_{Y_i}(y) = 1/2$, $y \in [-1, 1]$, $i = 1, 2$. Med lagen om total förväntan får vi

$$\begin{aligned} P(H) &= \int_0^1 \int_0^1 \frac{(2 - x_1)(2 - x_2)}{4} dx_1 dx_2 \\ &= \frac{9}{16} \end{aligned}$$

<u>Svar:</u> Den sökta sannolikheten är 9/16.