Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
ONSDAG 5 JUNI 2024 KL 08.00–13.00

Examinator: Björn-Olof Skytt, tel 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 (i svarsblanketten anges ordet Bonus). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet Bonus). Detta gäller endast på ordinarie tentan i mars 2024 och vid omtentamen i juni 2024. 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 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 den andra datorlaborationen får 3 bonuspoäng på del II på ordinarie tentamenstillfället och vid omtentamen i juni 2024.

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

A och B är två oberoende händelser där $P(A) = 0.5$ och $P(B) = 0.4$. Beräkna den betingade sannolikheten för att exakt en av händelserna A och B inträffar, då man vet att minst en av händelserna A och B har inträffat.

A: 0.5

B: 0.56

C: 0.625

D: 0.71

Uppgift 2

En stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ x^3 & \text{om } 0 \leq x \leq 1, \\ 1 & \text{om } x > 1. \end{cases}$$

Beräkna standardavvikelsen $D(2X)$.

A: 0.075

B: 0.15

C: 0.274

D: 0.387

Uppgift 3

Ur en vanlig kortlek, som består av 52 kort varav 13 i var och en av de 4 olika färgerna, drar man utan återläggning 4 kort på måfå. Beräkna sannolikheten att man då får ett kort i varje färg.

A: 0.03

B: 0.11

C: 0.17

D: 0.20

Uppgift 4

Den tvådimensionella stokastiska variabeln (X, Y) har den simultana sannolikhetsfunktionen:

$$p_{X,Y}(k, j) = \begin{cases} 0.2 & \text{om } k = j = 0, \\ 0.3 & \text{om } k = 0, j = 1, \\ 0.4 & \text{om } k = 1, j = 0, \\ 0.1 & \text{om } k = j = 1. \end{cases}$$

Beräkna korrelationskoefficienten $\rho(0.5X, 2Y)$

A: -0.10

B: -0.20

C: -0.41

D: -0.82

Uppgift 5

X och Y är oberoende stokastiska variabler.
$X \in N(2,4)$ och $Y \in N(3,3)$. Dvs. $E(X)=2$,$D(X)=4$,$E(Y)=3$,$D(Y)=3$.
Beräkna $P(X>Y)$.

A: 0.36

B: 0.39

C: 0.42

D: 0.44

Uppgift 6

Antalet samtal som inkommer till ett kontor anses vara en stokastisk variabel som är Poisson-fördelad med intensiteten 0.5 samtal per minut. Beräkna sannolikheten att det inkommer högst ett samtal till kontoret under en 5 min period.

A: 0.082

B: 0.205

C: 0.287

D: 0.503

Uppgift 7

Två undersökningar gjordes i två olika kommuner för att se hur många som tänker poströsta i det kommande EU-valet. I den ena kommunen svarade 14 av de 50 som svarade att de tänkte poströsta medan det i den andra kommunen var 36 av de 100 som svarade som uppgav att de tänkte poströsta. Låt p_1 stå för andelen röstberättigade i den ena kommunen som tänker poströsta och låt p_2 stå för andelen röstberättigade i den andra kommunen som tänker poströsta. Vi är intresserade av parametern $p_1 - p_2$. Bestäm medelfelet för skattningen av $p_1 - p_2$ om $p^*_{1_{obs}} - p^*_{2_{obs}} = \frac{14}{50} - \frac{36}{100}$.

A: 0.0796

B: 0.456

C: 0.657

D: 0.800

Uppgift 8

En stokastisk variabel X har följande fördelningsfunktion

$$F_X(x) = \begin{cases} 0; & x \leq 0 \\ \\ x^{\theta+1}; & 0 \leq x \leq 1 \\ \\ 1; & x \geq 1 \end{cases}$$

Vi har gjort tre oberoende försök och fått utfallen $x_1 = 0.70$, $x_2 = 0.80$, och $x_3 = 0.50$.
Bestäm Minsta-Kvadrat-skattningen av θ.

A: 0

B: 1

C: 2

D: 3

Uppgift 9

Vid tillverkning av skruvar får variationerna hos huvudenas diameter inte vara för stora. Man har mätt diametern hos 23 skruvar och beräknat stickprovets standardavvikelse $s = 0.031$ mm. Ange nedre gränsen för det tvåsidiga 95% konfidensintervallet för σ. Mätvärdena på diametrarna kan anses normalfördelade.

A: 0.019

B: 0.024

C: 0.0251

D: 0.027

Uppgift 10

Vi vill skatta skillnaden mellan väntevärdena μ_y och μ_x för två okända fördelningar och tar därför två slumpmässiga stickprov $x_1, x_2, ...x_n$ och $y_1, y_2, ...y_n$ av storlek $n = 30$. Stickprovsmedelvärdena blir $\bar{x} = 160$ och $\bar{y} = 220$, medan stickprovsstandardavvikelserna blir $s_x = 10$ och $s_y = 15$. Fördelningarna ser ut att vara tillräckligt symmetriska för att vi ska kunna anse att n är stort. Ange övre gränsen för det ensidigt uppåt begränsade konfidensintervallet för $\mu_y - \mu_x$ som har den approximativa konfidensgraden 95%

A: 65.41

B: 65.60

C: 66.45

D: 66.58

Uppgift 11

Antag att den stokastiska variabeln X är För-första-gången-fördelad. Dvs $X \in ffg(p)$. Vi ställer upp $H_0 : p = 0.1$. Vi vill testa nollhypotesen H_0 mot alternativet $H_1 : p = 0.4$ och förkastar H_0 till förmån för mothypotesen H_1 för små värden på x. Bestäm p-värdet om vi fått observationen $x = 2$.

A: 0.09

B: 0.19

C: 0.24

D: 0.64

Uppgift 12

Följande tabell ger observationer av elförbrukningen en viss dag i 3 samhällen.

Elförbrukningen i MWh	9	41	100
Invånarantal i tusental	1	5	12

Observationerna av elförbrukningen uppfattas som utfall av oberoende $N(\alpha + \beta x, \sigma)$-fördelade stokastiska variabler, där x är invånarantalet i tusental och α,β,och σ är okända parametrar. Invånarantalet i ett annat samhälle är 7000. Sätt upp en enkel linjär regressionsmodell för elförbrukningen som funktion av invånarantalet, skatta α och β, och skatta den förväntade elförbrukningen i detta samhälle utifrån dina skattningar av α och β.

A: 57.9

B: 58.3

C: 59.8

D: 60.1

Var god vänd!

Del II

Uppgift 13

En händelse A har sannoliketen p att inträffa i ett försök. Man utför försöket 125 gånger oberoende av varandra och bildar den stokastiska variabeln
$X =$ (Antalet gånger A inträffar) - (Antalet gånger A inte inträffar).

a) Beräkna $E(X)$ och $V(X)$ som funktion av p (7 p)

b) Beräkna $P(X \leq -111)$ då $p = 0.048$. Lämpliga och välmotiverade approximationer är tillåtna. (3 p)

Uppgift 14

Vid en läkarmottagning kallas 25 patienter till ett visst klockslag. Behandlingstiden för en patient betraktas som ett utfall av en exponentialfördelad stokastisk variabel med väntevärdet a minuter. Patienterna behandlas en i taget och olika patienters behandlingstider är oberoende. Av erfarenhet vet man att läkaren 1 gång på 10 klarar av att behandla de 25 patienterna inom en timme. Tolka denna uppgift som en sannolikhet och bestäm a. Rimliga och välmotiverade approximationer är tillåtna. (10 p)

Uppgift 15

Hos en urmakare har 480 klockor observerats. Nedanstående tabell visar antalet klockor, vars timvisare vid en viss tidpunkt står mellan 0 och 1, mellan 1 och 2 o.s.v.

Timvisarens läge	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Antal klockor	33	44	41	47	40	32	47	37	40	39	32	48

Kan hypotesen om likformig fördelning på intervallet 0-12 förkastas på risknivån 5%? Din slutsats skall givetvis vara ordentligt motiverad. (10 p)

Uppgift 16

Om vi har n st oberoende observationer $x_1, ..., x_n$ på en stokastisk variabel X med $E(X) = \mu$ så gäller att

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

är en rimlig skattning av μ. Antag att $X \in U(0, 2\mu)$ och bilda skattningen $\mu^*_{obs} = a_n max(x_1, ..., x_n)$.

a) Bestäm a_n så att μ^*_{obs} blir en väntevärdesriktig skattning av μ. (5 p)

b) Visa att μ^*_{obs} med a_n valt enligt a), är en effektivare skattning av μ än vad $\bar{x}$ är så snart $n \geq 2$.
För att du skall kunna få poäng på b) så måste a) vara helt rätt. (5 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG
TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
ONSDAG 5 JUNI 2024 KL 8.00–13.00.

Del I, Svar

1. D

2. D

3. B

4. C

5. C

6. C

7. A

8. B

9. B

10. A

11. B

12. B

Del I, Lösningsförslag

Uppgift 1

Vi söker

$$P(((A \cap B^*) \cup (B \cap A^*)) | (A \cup B)) = \frac{P((A \cap B^*) \cup (B \cap A^*))}{P(A \cup B)} = \text{disjunkta} = \frac{P(A \cap B^*) + P(B \cap A^*)}{P(A \cup B)} =$$

$$= \text{oberoende} = \frac{P(A)P(B^*) + P(A^*)P(B)}{P(A) + P(B) - P(A)P(B)} = \frac{0.5 \cdot 0.6 + 0.5 \cdot 0.4}{0.5 + 0.4 - 0.5 \cdot 0.4} = \frac{0.5}{0.7} = 0.71$$

Uppgift 2

En stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ x^3 & \text{om } 0 \leq x \leq 1, \\ 1 & \text{om } x > 1. \end{cases}$$

och således täthetsfunktionen

$$f_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ 3x^2 & \text{om } 0 \leq x \leq 1, \\ 0 & \text{om } x > 1. \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 3x^2 dx = [\frac{3x^4}{4}]_0^1 = \frac{3}{4}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \cdot 3x^2 dx = [\frac{3x^5}{5}]_0^1 = \frac{3}{5}$$

$$V(X) = E(X^2) - E^2(X) = \frac{3}{5} - \frac{3^2}{4^2} = \frac{3}{80} \Rightarrow V(2X) = \frac{4 \cdot 3}{80} \Rightarrow D(2X) = \sqrt{\frac{4 \cdot 3}{80}} = 0.387$$

Uppgift 3

Om vi tar hänsyn till ordningen är antalet möjliga utfall $52 \cdot 51 \cdot 50 \cdot 49$.
Antal gynnsamma utfall är om vi tar hänsyn till ordningen $52 \cdot 39 \cdot 26 \cdot 13$

$$\Rightarrow \text{ Sannolikheten för 4 olika färger blir } \frac{52 \cdot 51 \cdot 50 \cdot 49}{52 \cdot 39 \cdot 26 \cdot 13} = 0.11$$

Om vi inte tar hänsyn till ordningen får vi också att sannolikheten för 4 olika färger blir

$$\frac{13^4}{\binom{52}{4}} = 0.11$$

Uppgift 4

$$p_{X,Y}(k,j) = \begin{cases} 0.2 & \text{om } k = j = 0, \\ 0.3 & \text{om } k = 0, j = 1, \\ 0.4 & \text{om } k = 1, j = 0, \\ 0.1 & \text{om } k = j = 1. \end{cases}$$

Beräkna korrelationskoefficienten

$$\rho(0.5X, 2Y) = \frac{C(0.5X, 2Y)}{D(0.5X)D(2Y)} = \frac{0.5 \cdot 2C(X,Y)}{0.5 \cdot D(X) \cdot 2 \cdot D(Y)} = \rho(X,Y)$$

$$\rho(X,Y) = \frac{C(X,Y)}{D(X)D(Y)} = \frac{E(XY) - E(X)E(Y)}{D(X)D(Y)}$$

$$E(XY) = \sum_{allak,j} k \cdot j \cdot p_{X,Y}(k,j) = 0.1$$

$$E(X) = \sum_{allak,j} k \cdot p_{X,Y}(k,j) = 0.5$$

$$E(Y) = \sum_{allak,j} j \cdot p_{X,Y}(k,j) = 0.4$$

$\Rightarrow C(X,Y) = 0.1 - 0.5 \cdot 0.4 = -0.1$

$$E(X^2) = \sum_{allak,j} k^2 \cdot p_{X,Y}(k,j) = 0.5$$

$$E(Y^2) = \sum_{allak,j} j^2 \cdot p_{X,Y}(k,j) = 0.4$$

$V(X) = E(X^2) - E^2(X) = 0.5 - 0.25 = 0.25$
$V(Y) = E(Y^2) - E^2(Y) = 0.4 - 0.16 = 0.24$

$$\text{Alltså blir } \rho(0.5X, 2Y) = \rho(X,Y) = \frac{-0.1}{\sqrt{0.25}\sqrt{0.24}} = -0.41$$

Uppgift 5

X och Y är oberoende stokastiska variabler.
$X \in N(2,4)$ och $Y \in N(3,3)$. Dvs. $E(X) = 2$,$D(X) = 4$,$E(Y) = 3$,$D(Y) = 3$.

$P(X > Y) = P(X - Y > 0) = P(Z > 0)$
$E(Z) = 2 - 3 = -1$
X och Y är oberoende så $V(Z) = V(X - Y) = V(X) + (-1)^2 V(Y) = 16 + 9 = 25 \Rightarrow D(Z) = 5$

$Z \in N(-1, 5)$

$P(Z > 0) =$ Gör om till N(0,1) $= P(\frac{Z+1}{5} > \frac{0+1}{5}) = P(W > 0.2) = 1 - \Phi(0.2) = 1 - 0.58 = 0.42$

Uppgift 6

Låt X=antalet händelser under tiden t. Då är $X \in Po\left(\lambda t\right)$ Tiden mäts i minuter, så $t = 5$. Intensiteten per minut är $\lambda = \frac{1}{2}$. Därför gäller att $X \in Po\left(\frac{5}{2}\right)$

$$P\left(X \leq 1\right) = P\left(X = 0\right) + P\left(X = 1\right) = e^{-\frac{5}{2}} + \frac{5}{2}e^{-\frac{5}{2}} = \left(1 + \frac{5}{2}\right)e^{-\frac{5}{2}} = \frac{7}{2}e^{-\frac{5}{2}} = 0.287$$

Uppgift 7

Vi antar att antalet i den första kommunen som svarar att de tänker poströsta är en stokastisk variabel X där $X \in Bin(50, p_1)$ och att antalet i den andra kommunen som svarar att de tänker poströsta är en stokastisk variabel Y där $X \in Bin(100, p_2)$. Vi antar också att X och Y är oberoende.
Medelfelet av skattningen är $D^*_{obs}(p_1^* - p_2^*)$ (se F.S. §9.3).

$$V(p_1^* - p_2^*) = V\left(\frac{X}{50} - \frac{Y}{100}\right) = \text{p.g.a. oberoendet} = V\left(\frac{X}{50}\right) + (-1)^2 V\left(\frac{Y}{100}\right) = \frac{1}{50^2}V(X) + \frac{1}{100^2}V(Y)$$

$$= \frac{1}{50^2}50p_1(1-p_1) + \frac{1}{100^2}100p_2(1-p_2) \Rightarrow D(p_1^* - p_2^*) = \sqrt{\frac{1}{50}p_1(1-p_1) + \frac{1}{100}p_2(1-p_2)}$$

Då blir medelfelet

$$\left(\sqrt{\frac{1}{50}p_1(1-p_1) + \frac{1}{100}p_2(1-p_2)}\right)^*_{obs} = \sqrt{\frac{1}{50}p^*_{1_{obs}}(1-p^*_{1_{obs}}) + \frac{1}{100}p^*_{2_{obs}}(1-p^*_{2_{obs}})} =$$

$$= \sqrt{\frac{1}{50}0.28 \cdot 0.72 + \frac{1}{100}0.36 \cdot 0.64} = 0.0796$$

Uppgift 8

En stokastisk variabel X har följande fördelningsfunktion

$$F_X(x) = \begin{cases} 0; & x \leq 0 \\ \\ x^{\theta+1}; & 0 \leq x \leq 1 \\ \\ 1; & x \geq 1 \end{cases}$$

Vi har gjort tre oberoende försök och fått utfallen $x_1 = 0.70$, $x_2 = 0.80$, och $x_3 = 0.50$.
Bestäm Minsta-Kvadrat-skattningen av θ.

$$f_X(x) = \begin{cases} (\theta+1)x^{\theta}; & 0 \leq x \leq 1 \\ \\ 0; & annars \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x(\theta+1)x^{\theta} dx = [\frac{\theta+1}{\theta+2} x^{\theta+2}]_0^1 = \frac{\theta+1}{\theta+2}$$

$$Q = \sum_{j=1}^{3} (x_j - E(X))^2 = \left(0.7 - \frac{\theta+1}{\theta+2}\right)^2 + \left(0.8 - \frac{\theta+1}{\theta+2}\right)^2 + \left(0.5 - \frac{\theta+1}{\theta+2}\right)^2$$

$$\frac{dQ}{d\theta} = 2\left(0.7 - \frac{\theta+1}{\theta+2}\right)\left(\frac{-2}{(\theta+2)^2}\right) + 2\left(0.8 - \frac{\theta+1}{\theta+2}\right)\left(\frac{-2}{(\theta+2)^2}\right) + 2\left(0.5 - \frac{\theta+1}{\theta+2}\right)\left(\frac{-2}{(\theta+2)^2}\right) = 0$$

Detta leder till följande ekvation:

$$0.7 + 0.8 + 0.5 = 3\left(\frac{\theta+1}{\theta+2}\right) \Rightarrow 2\theta + 4 = 3\theta + 3 \Rightarrow \text{Minsta-Kvadrat-skatttningen av } \theta = 1$$

Uppgift 9

Då vi har ett oberoende, normalfördelat stickprov gäller enligt §11.1b) att

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

och enligt §12.4 att

$$f \cdot (\frac{\theta^*}{\theta})^2 \in \chi^2(f).$$

För att skapa det efterfrågade konfidensintervallet, kombinera §12.4 och §11.b.
Då fås konfidensintervallet för σ

$$\left(s\sqrt{\frac{(n-1)}{\chi^2_{0.025}(n-1)}}, s\sqrt{\frac{(n-1)}{\chi^2_{0.975}(n-1)}} \right).$$

Vilket med insatta numeriska värden ger

$$\left(0.031\sqrt{\frac{22}{\chi^2_{0.025}(22)}}, 0.031\sqrt{\frac{22}{\chi^2_{0.975}(22)}} \right) = (0.024, 0.044)$$

Uppgift 10

Vi ska bilda ett konfidensintervall för skillnaden mellan två stickprovs väntevärden då vi inte vet vilka fördelningar vi har. Vi kan då ta hjälp av §12.3 och bilda ett konfidensintervall med approximativ konfidensgrad eftersom vi kan anta att observationerna i varje stickprov är utfall från stokastiska variabler som är oberoende, likafördelade och många. Centrala Gränsvärdessatsen ger då att vi har ett $\theta^* = \mu_y^* - \mu_x^* = \bar{Y} - \bar{X}$ som är approximativt Normalfördelad och §12.3 kan då användas.
Vi behöver nu medelfelet $D^*_{obs} = D^*_{obs}(\theta^*)$

$$V(\theta^*) = V(\bar{Y} - \bar{X}) = \text{oberoende} = V(\bar{Y}) + V(\bar{X}) = \frac{\sigma_x^2}{30} + \frac{\sigma_y^2}{30}$$

$$D(\theta^*) = \sqrt{\frac{\sigma_x^2}{30} + \frac{\sigma_y^2}{30}} \Rightarrow D^*_{obs}(\theta^*) = \sqrt{\frac{s_x^2}{30} + \frac{s_y^2}{30}}$$

$$I_{\mu_y - \mu_x} = \bar{y} - \bar{x} \pm \lambda_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{30} + \frac{\sigma_y^2}{30}}$$

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

$$I_{\mu_y - \mu_x} = \left(-\infty, \bar{y} - \bar{x} + \lambda_\alpha \sqrt{\frac{\sigma_x^2}{30} + \frac{\sigma_y^2}{30}} \right)$$

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

$$\Rightarrow I_{\mu_y - \mu_x} = \left(-\infty, 220 - 160 + 1.6449\sqrt{\frac{10^2}{30} + \frac{15^2}{30}} \right) = (-\infty, 65.41)$$

Uppgift 11

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 $P(X \leq 2)$ om $X \in ffg(0.1)$
Då blir P-värdet $p_X(1) + p_X(2) = 0.1 + 0.9 \cdot 0.1 = 0.19$

Uppgift 12

$$\text{se § 13.1: } \beta^*_{obs} = \frac{\sum_{k=1}^{3}(x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^{3}(x_k - \bar{x})^2} = \text{ se § 13.3 } = \frac{\sum_{k=1}^{3} x_k y_k - 3\bar{x}\bar{y}}{\sum_{k=1}^{3} x_k^2 - 3\bar{x}^2}$$

$$\beta^*_{obs} = \frac{1 \cdot 9 + 5 \cdot 41 + 12 \cdot 100 - 3 \cdot 6 \cdot 50}{1^2 + 5^2 + 12^2 - 3 \cdot 6^2} = \frac{514}{62} = \frac{257}{31}$$

$$\alpha^*_{obs} = \text{ se § 13.1 } = \bar{y} - \beta^*_{obs}\bar{x} = 50 - \frac{257}{31} \cdot 6 = \frac{8}{31}$$

Den sökta elförbrukningen för $x = 7$ blir då

$$\frac{8}{31} + \frac{257}{31} \cdot 7 \approx 58.3$$

Del II, Lösningsförslag

Uppgift 13

Sätt Y = antalet gånger A inträffar. $Y \in Bin(125, p)$.

Sambandet mellan X och Y är: $X = Y - (125 - Y) = 2Y - 125$.

a) $E(Y) = 125p$, $V(Y) = 125p(1 - p)$

$E(X) = 2E(Y) - 125 = 250p - 125$

$V(X) = 2^2V(Y) = 500p(1 - p)$

b) $P(X \leq -111) = P(2Y - 125 \leq -111) = P(Y \leq 7)$

Men $Y \in Bin(125, p)$,där $p = 0.048 < 0.1$.

Alltså gäller att Y är approximativt Poissonfördelad:$Po(125 \cdot 0.048) = Po(6)$

Tabell 5 i F.S ger då $P(X \leq 7) = 0.744$

Uppgift 14

Låt för $k = 1, ...25$ den stokastiska variabeln X_k beteckna behandlingstiden i minuter för patient nr k.

$$\text{Sätt } Y = \sum_{k=1}^{25} X_k = \text{den sammanlagda behandligstiden}$$

Det är givet att $X_1, ..., X_{25}$ är oberoende och att alla $X_k \in Exp(1/a)$.

Detta ger $E(X_k) = a, V(X_k) = a^2$ för $k = 1, ...25$

$$E(Y) = \sum_{k=1}^{25} E(X_k) = 25a$$

$$V(Y) = \text{ oberoende } = \sum_{k=1}^{25} V(X_k) = 25a^2 \Rightarrow D(Y) = 5a$$

Vi har 25 st oberoende likafördelade stokastiska variabler som kan betraktas som många så Centrala Gränsvärdessatsen ger att $Y \in N(25a, 5a)$ approximativt. Att läkaren 1 gång på 10 klarar av att behandla 25 patienter inom en timme tolkar vi så att $P(Y \leq 60) = 0.1$
Vi gör om till $N(0, 1)$

$$P\left(\frac{Y - 25a}{5a} \leq \frac{60 - 25a}{5a}\right) = 0.1 \Rightarrow P\left(\frac{Y - 25a}{5a} \geq \frac{25a - 60}{5a}\right) = 0.1$$

$$\Rightarrow \frac{25a - 60}{5a} = \lambda_{0.10} \Rightarrow 25a - 60 = 5a\lambda_{0.10} \Rightarrow (25 - 5\lambda_{0.10})a = 60$$

$$a = \frac{60}{25 - 5 \cdot 1.2816} \approx 3.23$$

Uppgift 15

Här har vi test av given fördelning med nollhypotesen $H_0 : p_1 = \frac{1}{12}, p_2 = \frac{1}{12}, ..., p_{12} = \frac{1}{12}$.

Se § 14.3

Om $np_i \geq 5$ för alla i gäller att $Q \sim \chi^2$-fördelad, där

$$Q = \sum_{i=1}^{12} \frac{(x_i - np_i)^2}{np_i}$$

Här gäller att $np_i = 480 \cdot \frac{1}{12} = 40 \geq 5$ för alla i.

$$Q = \frac{(33-40)^2}{40} + \frac{(44-40)^2}{400} + ... + \frac{(48-40)^2}{40} = \frac{366}{40} = 9.15$$

Vi ska jämföra Q med $\chi^2_\alpha(r-1)$ där r = 12 i vårt fall och $\alpha = 0.05$.

$\chi^2_{0.05}(11) = 19.7$.

Om $Q < \chi^2_\alpha(r-1)$ så förkastas inte H_0 på risknivån α. Alltså kan H_0 inte förkastas på risknivån 5% i detta fall.

Uppgift 16

a) Vi ska bestämma a_n så att μ^*_{obs} blir en väntevärdesriktig skattning av μ.

Dvs. vi ska bestämma a_n så att $E(\mu^*) = \mu$

Sätt $Y_n = max(X_1, ..., X_n)$

$F_{Y_n}(y) = P(X_1 \le y \cap X_2 \le y \cap .... \cap X_n \le y) = [\text{oberoende}] = (F_{X_i}(y))^n$

$X_i \in U(0, 2\mu) \Rightarrow f_{X_i}(y) = \frac{1}{2\mu}$ om $0 \le y \le 2\mu$ och 0 annars

$F_{X_i}(y) = \ [0 \le y \le 2\mu] = \int_0^y \frac{1}{2\mu} dx = \frac{y}{2\mu}$

Detta leder till:

$$F_{Y_n}(y) = \begin{cases} 0 & \text{om } y < 0, \\ (\frac{y}{2\mu})^n & \text{om } 0 \le y \le 2\mu, \\ 1 & \text{om } y \ge 2\mu. \end{cases}$$

$$f_{Y_n}(y) = \begin{cases} \frac{n}{(2\mu)^n} y^{n-1} & \text{om } 0 \le y \le 2\mu, \\ 0 & \text{annars} \end{cases}$$

$E(Y_n) = \frac{n}{(2\mu)^n} \int_0^{2\mu} y \cdot y^{n-1} dy = 2\mu \frac{n}{n+1}$

$E(\mu^*) = E(a_n \cdot Y_n) = a_n \cdot 2\mu \frac{n}{n+1}$

$E(\mu^*) = \mu \Rightarrow a_n = \frac{n+1}{2n}$

b)

$$\text{Eftersom } E(\bar{X}) = \frac{1}{n} \sum_{k=1}^{n} E(X_k) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

så är $\bar{x}$ en väntevärdesriktig punktskattning av μ. I uppgift a) har vi bestämt a_n, så att även μ^*_{obs} är väntevärdesriktig.

Nu skall vi beräkna $V(\bar{X})$ och $V(\mu^*)$ och jämföra dessa med varandra.

$$V(\bar{X}) = \text{oberoende} = \frac{1}{n^2} \sum_{k=1}^{n} V(X_k) = \frac{1}{n^2} \cdot n \cdot \frac{(2\mu - 0)^2}{12} = \frac{\mu^2}{3n}$$

$$V(\mu^*) = V(a_n \cdot Y_n) = \left(\frac{n+1}{2n}\right)^2 V(Y_n)$$

$$E(Y_n^2) = \frac{n}{(2\mu)^n}\int_0^{2\mu} y^2 \cdot y^{n-1}dy = \frac{n}{(2\mu)^n}\left[\frac{y^{n+2}}{n+2}\right]_0^{2\mu} = (2\mu)^2\frac{n}{n+2}$$

$$V(Y_n) = E(Y_n^2) - E^2(Y_n) = (2\mu)^2\left(\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right) = (2\mu)^2\frac{n}{n+2} =$$

$$= (2\mu)^2\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2} = (2\mu)^2\frac{n^3+2n^2+n-n^3-2n^2}{(n+2)(n+1)^2} = (2\mu)^2\frac{n}{(n+2)(n+1)^2}$$

$$V(\mu^*) = \left(\frac{n+1}{2n}\right)V(Y_n) = \frac{\mu^2}{n(n+2)}$$

$$\text{Om } n = 1 \text{ blir } V(\mu^*) = V(\bar{X}) = \frac{\mu^2}{3}$$

$$\text{Om } n \geq 2 \text{ gäller att } \frac{V(\bar{X})}{V(\mu^*)} = \frac{\frac{\mu^2}{3n}}{\frac{\mu^2}{n(n+2)}} = \frac{n+2}{3} > 1 \text{ om } n \geq 2$$

Alltså gäller att $V(\mu^*) < V(\bar{X})$ om $n \geq 2$, vilket betyder att μ^*_{obs} är en effektivare skattning än $\bar{x}$.