



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1923 SANNOLIKHETSTEORI OCH STATISTIK, SAMT TENTAMEN I SF1935 SANNOLIKHETSTEORI OCH STATISTIK MED TILLÄMPNING INOM MASKININLÄRNING
TISDAG 18 AUGUSTI 2026 KL 8.00–13.00.

Kursledare: 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 och varje korrekt svar ger 1 poäng. På denna del ska endast svar anges i form av val av ett av de möjliga svarsalternativen på de uppgifter där svarsalternativ finns, annars skall svaret ges som ett numeriskt värde med minst *fyra* värdesiffrors noggrannhet! Svaren ska anges på svarsblanketten (utdelas vid tentamen). Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1–3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges då ordet "Bonus"). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges då ordet "Bonus"). Dessa tillgodoräknanden gäller endast för ordinarie tentamen i maj 2026 och vid omtentamen i augusti 2026. 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 ska resonemang och uträkningar vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar ska förklaras och definieras samt numeriska svar ska anges med minst *fyra* värdesiffrors noggrannhet. Studenter som är godkända på den andra datorlaborationen får dessutom tre bonuspoäng på del II. Dessa bonuspoäng gäller endast för ordinarie tentamen i maj 2026 och vid omtentamen i augusti 2026.

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

När ett företag skickar fraktgodis till en återförsäljare sker detta antingen med buss, tåg eller flyg. 50% fraktas med buss, 30% med tåg och 20% med flyg. Andelen transportskadat gods är 5% med buss, 10% med tåg och 2% med flyg. Om man mottar ett transportskadat gods hur stor är sannolikheten att det har skickats med flyg?

Uppgift 2

Den stokastiska variabeln X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ x^2/4 & \text{om } 0 \leq x < 2, \\ 1 & \text{om } x \geq 2. \end{cases}$$

En ny stokastisk variabel Y bildas genom $Y = -2X + 1$.

Bestäm $P(|Y| \geq 1)$.

A: 0.75

B: 0.50

C: 0.25

D: 0.0625

Uppgift 3

Den tvådimensionella s.v. (X, Y) antas vara likformigt fördelad på kvadraten $K = \{(x, y) : 0 \leq x \leq 2, 0 \leq y \leq 2\}$. Beräkna $E(X^3 Y^3)$.

A: 16

B: 4

C: 2

D: 1

Uppgift 4

Låt X och Y vara oberoende stokastiska variabler, sådana att $X \in \text{Bin}(3, 0.2)$ och $Y \in \text{Bin}(5, 0.2)$. Bestäm $P(X + Y = 4)$.

A: 0.046

B: 0.954

C: 0.013

D: 0.99

Uppgift 5

Slumpvariabeln X har exponentialfördelning med väntevärdet 200 timmar. Beräkna medianen i fördelningen.

Uppgift 6

En normalfördelad stokastisk variabel X med väntevärde $E(X) = -2$, har egenskapen att $P(X \leq 0) = 0.95$. Låt $Y = 2X - 1$.

Beräkna $D(Y)$.

A: 2.864

B: 2.432

C: 1.432

D: 1.216

Uppgift 7

Antag att X är fördelad enligt $Po(\mu)$ medan Y är fördelad enligt $Po(5\mu)$.

Bestäm Minsta-Kvadrat-skattningen för μ givet $x = 10$ och $y = 35$.

A: 22.5

B: 8.5

C: 7.583

D: 7.0

Uppgift 8

Antag att $x_1, x_2, \dots, x_5$ är ett observerat stickprov från $N(\mu, \sigma)$, där $\sigma = 2.3$. För att testa $H_0 : \mu = 125$ mot $H_1 : \mu < 125$ ville man använda beslutsregeln: förkasta H_0 om intervallet I_μ inte innehåller talet 125. Signifikansnivån ska bli 10 %.

Vilket av följande konfidensintervall är det korrekta då ?

A: $I_\mu = (\bar{x} - a, \infty)$, där $a = 1.32$

B: $I_\mu = (\bar{x} - a, \infty)$, där $a = 1.69$

C: $I_\mu = (-\infty, \bar{x} + a)$, där $a = 1.32$

D: $I_\mu = (-\infty, \bar{x} + a)$, där $a = 1.69$

Uppgift 9

Låt $X \in Po(\mu)$. Vi gör 25 oberoende observationer på X och får medelvärde till $\bar{x} = 880$. Ange nedre gränsen för det tvåsidiga konfidensintervallet för μ som har den approximativa konfidensgraden 99 %.

- A: 866.2
- B: 864.7
- C: 845.8
- D: 798.1

Uppgift 10

En forskare har gjort tio försök som anses vara oberoende av varandra där sannolikheten för lyckat försök är p . Låt den stokastiska variabeln X stå för antalet lyckade försök. Forskaren önskar pröva $H_0 : p = 1/2$ mot $H_1 : p > 1/2$. Man förkastar H_0 till förmån för H_1 om man får ett utfall $x \geq 8$. Resultatet av de tio försöken var åtta lyckade försök.

Bestäm testets styrka för alternativet $H_1 : p = 3/4$.

- A: 0.0547
- B: 0.9996
- C: 0.4744
- D: 0.5256

Uppgift 11

(Från en riktig studie år 2026) Vid en undersökning rörande CTFYS' deltagande i klassrum intervjuades 110 slumpmässigt utvalda studerande. Följande svarsfrekvenser erhöles:

	Inga eller några	Ungefär hälften	Majoriteten	Nästan alla
Jag deltar på föreläsningar	17	23	33	37
Jag deltar på övningar	40	27	23	20

Verkar det vara fler som går på föreläsningar än på övningar? För att svara på denna fråga utför man ett homogenitetstest, där teststorheten Q fås till $Q = 16.457$. Vilket av följande påståenden är sant?

- A: P-värde < 0.01 , och därmed verkar det vara fler som går på föreläsningar än på övningar på nivå 1 %
- B: P-värde < 0.01 , och därmed verkar det inte vara fler som går på föreläsningar än på övningar på nivå 1 %
- C: P-värde > 0.01 , och därmed verkar det vara fler som går på föreläsningar än på övningar på nivå 1 %
- D: P-värde > 0.01 , och därmed verkar det inte vara fler som går på föreläsningar än på övningar på nivå 1 %

Uppgift 12

Vid enkel linjär regression gäller följande samband

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

där man observerat paren (x_i, y_i) , $i = 1, \dots, n$ och ε_i betecknar de slumpmässiga felen. Man önskar testa nollhypotesen $H_0 : \beta = 0$ mot $H_1 : \beta \neq 0$ och får p -värdet till 0.072. Vilket av följande påståenden är sant?

- A: Vi förkastar H_0 och kan därmed förkasta att y_i :na beror linjärt av x_i :na på nivån 10%.
- B: Vi förkastar inte H_0 och kan därmed varken bekräfta eller definitivt utesluta ett linjärt beroende utifrån detta test på nivån 10%.
- C: Vi förkastar H_0 och kan därmed inte förkasta att y_i :na beror linjärt av x_i :na på nivån 5%.
- D: Vi förkastar inte H_0 och kan därmed varken bekräfta eller definitivt utesluta ett linjärt beroende utifrån detta test på nivån 5%.

Var god vänd!

Del II

Uppgift 13

Ett radioaktivt preparat avger impulser (t.ex. α -partiklar), där antalet impulser på intervallet Δt är $Po(\lambda\Delta t)$ -fördelade och där λ = intensiteten. Dvs. om $X(t_1, t_2)$ = antal impulser mellan tiden t_1 och tiden t_2 så gäller att $X(t_1, t_2) \in Po(\lambda(t_2 - t_1))$. Vi läser av mätaren först vid tiden $t = 1$ och konstaterar att preparatet avgett 2 impulser mellan tiden $t = 0$ och tiden $t = 1$. Sedan läser vi av mätaren vid tiden $t = 5$ och konstaterar att preparatet avgett 5 impulser mellan tiden $t = 0$ och tiden $t = 5$. Vad är då den betingade sannolikheten att preparatet avgett 3 impulser mellan tiden $t = 0$ och tiden $t = 2$?

Ledning: antal impulser på disjunkta intervall är oberoende.

(10 p)

Uppgift 14

Antag att den stokastiska variabeln $X \in U(0, 15)$ och att den stokastiska variabeln $Y \in U(0, 20)$. Antag vidare att X och Y är oberoende. Sätt nu $Z = \min(X, Y)$ och $W = \max(X, Y)$.

a) Ange fördelningsfunktionen $F_Z(z)$. (3 p)

b) Ange medianen $\tilde{z}$ och väntevärdet $E(Z)$. (2 p)

c) Ange fördelningsfunktionen $F_W(w)$. (3 p)

d) Ange medianen $\tilde{w}$ och väntevärdet $E(W)$. (2 p)

Uppgift 15

Tiden X mellan nervimpulser antas bestå dels av en konstant tid a , dels en stokastisk tid som är exponentialfördelad med väntevärde μ . Dvs. vi har täthetsfunktionen

$$f_X(x) = \begin{cases} \frac{1}{\mu} e^{-(x-a)/\mu} & \text{om } x \geq a \\ 0 & \text{om } x < a \end{cases}$$

Härled Maximum-Likelihood-skattningarna av μ och a då man har ett stickprov $x_1, x_2, \dots, x_n$ och beräkna skattningarna då man har registrerat följande tider mellan nervimpulser (enhet millisek).

7, 9, 12, 8, 14, 9, 17, 8, 13, 10

(10 p)

Uppgift 16

Antalet kolonier av koliforma bakterier i en milliliter vatten kan antas vara Poissonfördelat och oberoende av antalet kolonier i andra vattenprov/milliliter.

Vid två badställen tas två vattenprover:

	Provstorlek (ml)	Antal kolonier
Ihreviken	20	54
Gnisvärd	50	185

Testa om det genomsnittliga antalet kolonier per ml skiljer sig åt mellan badplatserna genom att bilda ett välmotiverat konfidensintervall. Använd en risknivå som är approximativt 1%. Dina hypoteser skall vara ordentligt angivna och din slutsats skall vara tydlig och välmotiverad .

(10 p)

Lycka till!



Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1923 SANNOLIKHETSTEORI OCH STATISTIK, SAMT
TENTAMEN I SF1935 SANNOLIKHETSTEORI OCH STATISTIK MED TILLÄMPNING IN-
OM MASKININLÄRNING
TISDAG 18 AUGUSTI 2026 KL 8.00–13.00.

Del I, Svar.

1. 0.0678 (alt 0.0677966)
2. A
3. B
4. A
5. 138.6 (alt ett godtyckligt tal från intervallet $[138, 139]$)
6. B
7. C
8. C
9. B
10. D
11. A
12. D

Del I, Lösningsförslag.**Uppgift 1**

Låt B, T, F vara händelserna att godset skickas med buss, tåg respektive flyg, och S att godset är transportskadat. Lagen om total sannolikhet ger:

$$\begin{aligned} P(S) &= P(S | B)P(B) + P(S | T)P(T) + P(S | F)P(F) = \\ &= 0.05 \cdot 0.5 + 0.10 \cdot 0.3 + 0.02 \cdot 0.2 = 0.025 + 0.03 + 0.004 = 0.059. \end{aligned}$$

Bayes sats ger:

$$P(F | S) = \frac{P(S | F)P(F)}{P(S)} = \frac{0.02 \cdot 0.2}{0.059} = \frac{0.004}{0.059} \approx 0.0678$$

Svar: ≈ 0.0678

Uppgift 2

Från F_X ser vi att X har stöd på $[0, 2]$ med $F_X(x) = x^2/4$ där $0 \leq x < 2$.
 $Y = -2X + 1$. Vi undersöker $P(|Y| \geq 1) = P(Y \geq 1) + P(Y \leq -1)$.

$$Y \geq 1 \iff -2X + 1 \geq 1 \iff X \leq 0.$$

Eftersom $X \geq 0$ med kontinuerlig fördelning är $P(X \leq 0) = F_X(0) = 0$.

$$Y \leq -1 \iff -2X + 1 \leq -1 \iff X \geq 1.$$

$$P(X \geq 1) = 1 - F_X(1) = 1 - \frac{1^2}{4} = 1 - 0.25 = 0.75.$$

Alltså är $P(|Y| \geq 1) = 0 + 0.75 = 0.75$

Svar: A.

Uppgift 3

Eftersom (X, Y) är likformigt fördelad på kvadraten $[0, 2] \times [0, 2]$ är X och Y oberoende, var och en $U(0, 2)$ -fördelad.

$$E(X^3) = \int_0^2 x^3 \cdot \frac{1}{2} dx = \frac{1}{2} \left[\frac{x^4}{4} \right]_0^2 = \frac{1}{2} \cdot 4 = 2.$$

Av symmetriskäl är även $E(Y^3) = 2$. Eftersom X, Y är oberoende:

$$E(X^3 Y^3) = E(X^3)E(Y^3) = 2 \cdot 2 = 4.$$

Svar: B.

Uppgift 4

Eftersom X och Y är oberoende binomialfördelade med samma sannolikhet $p = 0.2$ gäller att

$$X + Y \in \text{Bin}(3 + 5, 0.2) = \text{Bin}(8, 0.2).$$

$$P(X + Y = 4) = \binom{8}{4} 0.2^4 \cdot 0.8^4 = 70 \cdot 0.0016 \cdot 0.4096 \approx 0.04588 \approx 0.046.$$

(Kan också verifieras genom faltning av de två binomialfördelningarna direkt.)

Svar: A.

Uppgift 5

$X \in \text{Exp}(\lambda)$ med $E(X) = 200 \Rightarrow \lambda = 1/200$. Medianen m löser $F_X(m) = 0.5$:

$$1 - e^{-m/200} = 0.5 \implies m = 200 \ln 2 \approx 138.6$$

Svar: ≈ 138.6 timmar.

Uppgift 6

Låt $\sigma = D(X)$. Standardisering av $P(X \leq 0) = 0.95$ med $E(X) = -2$ ger

$$P\left(\frac{X - (-2)}{\sigma} \leq \frac{0 - (-2)}{\sigma}\right) = \Phi\left(\frac{2}{\sigma}\right) = 0.95 \implies \frac{2}{\sigma} = \lambda_{0.05} = 1.6449,$$

där $\lambda_{0.05}$ är den övre 5 %-kvantilen i $N(0, 1)$ (dvs. $\Phi(\lambda_{0.05}) = 0.95$). Alltså är

$$\sigma = \frac{2}{1.6449} \approx 1.2159.$$

Eftersom $Y = 2X - 1$ är en linjär transformation gäller $D(Y) = |2| \cdot D(X) = 2\sigma \approx 2 \cdot 1.2159 = 2.432$

Svar: B.

Uppgift 7

Eftersom $X \in \text{Po}(\mu)$ och $Y \in \text{Po}(5\mu)$ har vi

$$E(X) = \mu, \quad V(X) = \mu, \quad E(Y) = 5\mu, \quad V(Y) = 5\mu.$$

Eftersom X och Y har *olika varians* ska minsta-kvadrat-metoden vikta de två avvikelserna med den inversa variansen, så att observationer med större spridning får mindre inflytande. Detta ger kostnadsfunktionen

$$Q(\mu) = \frac{(x - \mu)^2}{V(X)} + \frac{(y - 5\mu)^2}{V(Y)} = \frac{(x - \mu)^2}{\mu} + \frac{(y - 5\mu)^2}{5\mu}.$$

Vi deriverar med avseende på μ (kvotregeln på respektive term):

$$\frac{d}{d\mu} \left[\frac{(x - \mu)^2}{\mu} \right] = 1 - \frac{x^2}{\mu^2}, \quad \frac{d}{d\mu} \left[\frac{(y - 5\mu)^2}{5\mu} \right] = 5 - \frac{y^2}{5\mu^2}.$$

Sätter vi summan till noll:

$$\frac{dQ}{d\mu} = 6 - \frac{x^2}{\mu^2} - \frac{y^2}{5\mu^2} = 0 \implies 6\mu^2 = x^2 + \frac{y^2}{5} \implies \mu_{obs}^* = \sqrt{\frac{x^2 + y^2/5}{6}}.$$

(Andraderivatet är positivt i denna punkt, så det är fråga om ett minimum.)

Med $x = 10$, $y = 35$:

$$\mu_{obs}^* = \sqrt{\frac{10^2 + 35^2/5}{6}} = \sqrt{\frac{100 + 245}{6}} = \sqrt{\frac{345}{6}} = \sqrt{57.5} \approx 7.583$$

Svar: C.

Uppgift 8

Eftersom $H_1 : \mu < 125$ är ensidig åt vänster ska vi förkasta H_0 för små värden på $\bar{x}$, vilket motsvarar ett ensidigt *uppåt begränsat* konfidensintervall av formen $I_\mu = (-\infty, \bar{x} + a]$ (vi förkastar H_0 när 125 ligger *utanför*, dvs. över, intervallet). Detta utesluter A och B (som har fel riktning, lämpliga för $H_1 : \mu > 125$).

För ett ensidigt test med signifikansnivå $\alpha = 0.10$ används den övre 10 %-kvantilen $\lambda_{0.10} = 1.2816$ (ej $\lambda_{0.05}$, som skulle motsvara ett tvåsidigt test eller fel signifikansnivå):

$$a = \lambda_{0.10} \cdot \frac{\sigma}{\sqrt{n}} = 1.2816 \cdot \frac{2.3}{\sqrt{5}} \approx 1.2816 \cdot 1.0288 \approx 1.318 \approx 1.32.$$

Svar: C.

Uppgift 9

$X \in \text{Po}(\mu)$, $n = 25$, $\bar{x}_{obs} = 880 \geq 15$, så normalapproximation kan användas: $\bar{X} \approx N(\mu, \sqrt{\mu/n})$.

$$I_\mu = \bar{x} \pm \lambda_{0.005} \sqrt{\frac{\bar{x}}{n}} = 880 \pm 2.5758 \sqrt{\frac{880}{25}} = 880 \pm 2.5758 \cdot 5.933 \approx 880 \pm 15.28.$$

Nedre gränsen är alltså $880 - 15.28 \approx 864.7$.

Svar: B.

Uppgift 10

Testets styrka är $h(3/4) = P(X \geq 8 \mid p = 3/4)$, där $X \in \text{Bin}(10, 0.75)$ (oberoende av det faktiska utfallet $x = 8$, eftersom styrkan är en egenskap hos testet för ett givet H_1 , inte hos ett enskilt observerat utfall).

$$h(3/4) = P(X \geq 8) = \sum_{k=8}^{10} \binom{10}{k} 0.75^k 0.25^{10-k} \approx 0.2816 + 0.1877 + 0.0563 \approx 0.5256,$$

alt använda Tabell 6 i FS, dvs

$$h(3/4) = P(X \geq 8) = P(Y \leq 2) = |\text{Tabell 6 för } Y \in \text{Bin}(10, 0.25)| \approx 0.5256$$

Svar: D.

Uppgift 11

Vi har ett homogenitetstest med en 2×4 -tabell (2 grupper: föreläsningar/övningar; 4 kategorier), så antalet frihetsgrader är $(2 - 1)(4 - 1) = 3$.

$$\chi_{0.01}^2(3) \approx 11.34.$$

Eftersom $Q = 16.457 > 11.34$ är P -värdet < 0.01 , så H_0 (homogenitet, dvs. samma svarsfördelning i båda grupperna) förkastas på nivån 1 %.

En jämförelse av rådata visar att kategorierna "Majoriteten" och "Nästan alla" är klart överrepresenterade för föreläsningar ($33+37=70$ av 110) jämfört med övningar ($23+20=43$ av 110), medan "Inga eller några" är klart vanligare för övningar (40 mot 17). Detta stöder slutsatsen att fler verkar gå på föreläsningar än på övningar.

Svar: A.

Mindre metodologisk anmärkning: de två raderna (föreläsningsnärvaro och övningsnärvaro) verkar komma från samma 110 studenter som besvarar två relaterade frågor, snarare än två oberoende urval. Ett standardiserat chi-kvadrattest för homogenitet/oberoende antar formellt oberoende urval; med parade/upprepade svar från samma individer skulle ett annat test (t.ex., McNemars-test för 2×2 tabeller eller ett Stuart-Maxwells test/marginal homogenitetstest ananns) vara det metodologiskt exakta valet. Detta är en vanlig förenkling i enkätbaserade urvalproblem och inte nödvändigtvis ett fel man behöver åtgärda.

Uppgift 12

P -värdet $0.072 > 0.05 = \alpha$, så vi förkastar *inte* $H_0 : \beta = 0$ på nivån 5 %. Att inte kunna förkasta H_0 innebär *inte* att vi bevisat att $\beta = 0$ (dvs. att inget linjärt samband finns) – vi saknar helt enkelt tillräckliga statistiska belägg för att med säkerhet fastställa ett linjärt samband. Vi kan således varken bekräfta eller definitivt utesluta ett linjärt beroende utifrån detta test.

Alternativet C är felaktigt redan i sitt första påstående (att H_0 skulle förkastas, vilket är fel då P -värdet > 0.05). Alternativet B är *inte* giltigt heller, ty P -värdet $= 0.072 < 0.1$, dvs vi förkastar $H_0 : \beta = 0$ på nivån 10%. Alternativet A är felaktigt eftersom att förkasta $H_0 : \beta = 0$ innebär att vi inte kan förkasta ett linjärt beroende utifrån detta test.

Svar: D.

Del II, Lösningsförslag.**Uppgift 13**

Vi börjar med att införa stokastiska variabler $X(t_1, t_2)$ där $X(t_1, t_2)$ = antal impulser mellan tiden t_1 och tiden t_2 .

Vi söker då $P(X(0, 2) = 3 | X(0, 1) = 2 \cap X(0, 5) = 5)$

$$\begin{aligned} P(X(0, 2) = 3 | X(0, 1) = 2 \cap X(0, 5) = 5) &= \frac{P(X(0, 2) = 3 \cap X(0, 1) = 2 \cap X(0, 5) = 5)}{P(X(0, 1) = 2 \cap X(0, 5) = 5)} = \\ &= \frac{P(X(0, 1) = 2 \cap X(1, 2) = 1 \cap X(2, 5) = 2)}{P(X(0, 1) = 2 \cap X(1, 5) = 3)} = [\text{antal impulser på disjunkta intervall är oberoende}] = \\ &= \frac{P(X(0, 1) = 2)P(X(1, 2) = 1)P(X(2, 5) = 2)}{P(X(0, 1) = 2)P(X(1, 5) = 3)} = \frac{P(X(1, 2) = 1)P(X(2, 5) = 2)}{P(X(1, 5) = 3)} = \\ &= \frac{\frac{(\lambda \cdot 1)^1}{1!} e^{-\lambda \cdot 1} \cdot \frac{(\lambda \cdot 3)^2}{2!} e^{-\lambda \cdot 3}}{\frac{(\lambda \cdot 4)^3}{3!} e^{-\lambda \cdot 4}} = \frac{\frac{9}{2}}{\frac{64}{6}} = \frac{27}{64} = 0.421875 \approx 0.4219 \end{aligned}$$

En alternativ lösning.

Man kan också resonera direkt med den välkända egenskapen att, givet det totala antalet punkter $N(t)$ i ett intervall $(0, t]$ för en Poissonprocess, är dessa punkter oberoende och likformigt fördelade över intervallet. Vilket medför att antalet punkter i en delmängd är binomialfördelat med sannolikhetsparameter p som är lika med kvoten mellan del-längden och hela längden.

Givet $N(5) - N(1) = 3$ (dvs. 3 punkter i intervallet $(1, 5]$ av längd 4), är antalet av dessa punkter som hamnar i delintervallet $(1, 2]$ (längd 1) $\text{Bin}(3, \frac{1}{4})$ -fördelat.

Vi söker sannolikheten att exakt 1 av de 3 punkterna hamnar i $(1, 2]$:

$$P(N(2) - N(1) = 1 | N(5) - N(1) = 3) = \binom{3}{1} \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^2 = 3 \cdot \frac{1}{4} \cdot \frac{9}{16} = \frac{27}{64} \approx 0.4219,$$

vilket ger samma svar som ovan (och är oberoende av att vi också villkorar på $N(1) = 2$, eftersom $N(1)$ är oberoende av ökningarna efter tidpunkt 1). Dvs, slutligen

$P(N(2) = 3 N(1) = 2, N(5) = 5) = P(N(2) - N(1) = 1 N(5) - N(1) = 3) = \frac{27}{64} \approx 0.4219$
--

Uppgift 14

Antag att den stokastiska variabeln $X \in U(0, 15)$ och att den stokastiska variabeln $Y \in U(0, 20)$.
Antag vidare att X och Y är oberoende. Sätt nu $Z = \min(X, Y)$ och $W = \max(X, Y)$.

a) Ange fördelningsfunktionen $F_Z(z)$.

Eftersom det är minimum det gäller räcker det att titta på området $0 \leq z \leq 15$.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(\min(X, Y) \leq z) = 1 - P(\min(X, Y) \geq z) = [X \text{ och } Y \text{ är oberoende}] = \\ &= 1 - P(X \geq z)P(Y \geq z) = 1 - [1 - P(X \leq z)][1 - P(Y \leq z)] = 1 - [1 - P(X \leq z)][1 - P(Y \leq z)] = \\ &= 1 - [1 - \frac{z}{15}][1 - \frac{z}{20}] = \frac{z}{15} + \frac{z}{20} - \frac{z^2}{15 \cdot 20} \end{aligned}$$

b) Medianen $\tilde{z}$ definieras av att $F_Z(\tilde{z}) = 0.5$

$$\begin{aligned} \frac{\tilde{z}}{15} + \frac{\tilde{z}}{20} - \frac{\tilde{z}^2}{15 \cdot 20} &= 0.5 \Rightarrow \frac{35\tilde{z}}{15 \cdot 20} - \frac{\tilde{z}^2}{15 \cdot 20} = 0.5 \Rightarrow \tilde{z}^2 - 35\tilde{z} + 150 = 0 \Rightarrow \\ \Rightarrow \tilde{z} &= \frac{35}{2} \pm \sqrt{\frac{35^2}{4} - 150} \Rightarrow \tilde{z} = 17.5 \pm 12.5 \end{aligned}$$

Dvs. medianen $\tilde{z} = 5$.

Väntevärdet $E(Z) = \int_0^{15} z f_Z(z) dz$

$$\begin{aligned} f_Z(z) &= \frac{dF_Z}{dz} = \frac{1}{15} + \frac{1}{20} - \frac{2z}{15 \cdot 20} \\ E(Z) &= \int_0^{15} \left[\frac{z}{15} + \frac{z}{20} - \frac{2z^2}{15 \cdot 20} \right] dz = \left[\frac{z^2}{30} + \frac{z^2}{40} - \frac{2z^3}{3 \cdot 15 \cdot 20} \right]_0^{15} = 5.625 \end{aligned}$$

c) Ange fördelningsfunktionen $F_W(w)$. Eftersom det är maximum det gäller tittar vi dels på området $0 \leq w \leq 15$, dels på området $15 \leq w \leq 20$.

$0 \leq w \leq 15$:

$$\begin{aligned} F_W(w) &= P(W \leq z) = P(\max(X, Y) \leq w) = [X \text{ och } Y \text{ är oberoende}] = P(X \leq w)P(Y \leq w) = \\ &= \frac{w}{15} \cdot \frac{w}{20} = \frac{w^2}{300} \\ F_W(15) &= \frac{3}{4} \end{aligned}$$

Eftersom maxvärdet på området $15 \leq w \leq 20$ bara beror av Y och vi vet att $Y \in U(0, 20)$ och att $F_W(20) = 1$, så gäller här att

$$F_W(w) = \frac{3}{4} + \frac{w - 15}{20}$$

Dvs.

$$F_W(w) = \begin{cases} 0 & \text{om } w \leq 0 \\ \frac{w^2}{300} & \text{om } 0 \leq w \leq 15 \\ \frac{3}{4} + \frac{w-15}{20} & \text{om } 15 \leq w \leq 20 \\ 1 & \text{om } w \geq 20 \end{cases}$$

d) När vi ska hitta medianen $\tilde{w}$ räcker det att titta på området $0 \leq w \leq 15$ eftersom $F_W(15) = \frac{3}{4}$

$$F_W(\tilde{w}) = 0.5 \Rightarrow \frac{\tilde{w}^2}{300} = 0.5 \Rightarrow \tilde{w}^2 = 150 \Rightarrow \tilde{w} = \sqrt{150} = 12.25$$

Dvs. medianen $\tilde{w} = 12.25$.

$$f_W(w) = \frac{dF_W}{dw} = [\text{om } 0 \leq w \leq 15] = \frac{w}{150}$$

$$f_W(w) = \frac{dF_W}{dw} = [\text{om } 15 \leq w \leq 20] = \frac{1}{20}$$

$$\Rightarrow \text{Väntevärdet } E(W) = \int_0^{15} w \frac{w}{150} dw + \int_{15}^{20} w \frac{1}{20} dw = \left[\frac{w^3}{450} \right]_0^{15} + \left[\frac{w^2}{40} \right]_{15}^{20} = \frac{15}{2} + 10 - \frac{15^2}{40} = 11.875$$

Uppgift 15

Vi söker det μ och det a som maximerar Likelihoodfunktionen $L(\mu, a)$, se § 9.1.

$$L(\mu, a) = \frac{1}{\mu} e^{-(x_1-a)/\mu} \cdot \frac{1}{\mu} e^{-(x_2-a)/\mu} \dots \cdot \frac{1}{\mu} e^{-(x_n-a)/\mu} = \frac{1}{\mu^n} e^{-\Sigma(x_i-a)/\mu}$$

Där vi hittar maximum för $\ln L(\mu, a)$ hittar vi maximum för $L(\mu, a)$

$$\ln L(\mu, a) = \ln \mu^{-n} - \frac{\Sigma(x_i - a)}{\mu} = -n \ln \mu - \frac{\Sigma(x_i - a)}{\mu}$$

$$\frac{\partial \ln L(\mu, a)}{\partial \mu} = \frac{-n}{\mu} + \frac{\Sigma(x_i - a)}{\mu^2} = 0 \Rightarrow \mu_{obs_{ML}}^* = \frac{\Sigma(x_i - a_{obs}^*)}{n}$$

Vi kollar om andraderivatans är < 0 .

$$\frac{\partial^2 \ln L(\mu, a)}{\partial^2 \mu} = \frac{n}{\mu^2} - \frac{2\Sigma(x_i - a)}{\mu^3}$$

$$[\text{vi förkortar bort } \mu^2 \text{ i nämnarna och sätter in skattningen av } \mu = \mu_{obs_{ML}}^* = \frac{\Sigma(x_i - a_{obs}^*)}{n}] \Rightarrow n - 2n < 0$$

$$\text{Alltså har vi maximum för } \mu = \frac{\Sigma(x_i - a_{obs}^*)}{n}$$

Det a som maximerar $L(\mu, a) = \frac{1}{\mu^n} e^{-\Sigma(x_i-a)/\mu}$ är det a som minimerar $\Sigma(x_i - a) = \Sigma x_i - na$ med bivillkoret enligt täthetsfunktionen att det för alla x_i gäller att $a \leq x_i$.

Alltså är ML-skattningen av $a = \min x_i$.

Dvs. $a_{obs_{ML}}^* = \min x_i = 7$ och $\mu_{obs_{ML}}^* = \frac{\Sigma(x_i - a_{obs_{ML}}^*)}{n} = \frac{\Sigma x_i}{10} - 7 = 10.7 - 7 = 3.7$

Uppgift 16

Vattenprovet från Ihreviken består av $n_1 = 20$ ml vatten. Antalet kolonier i dessa vattenprov, $x = 54$, kan ses som ett utfall av $X = X_1 + \dots + X_{n_1}$ där $X_1, \dots, X_{n_1}$ är oberoende och $\text{Po}(\mu_1)$ -fördelade. Alltså är X $\text{Po}(n_1\mu_1)$ -fördelad. På samma sätt består vattenprovet från Gnisvård av $n_2 = 50$ ml och antalet kolonier $y = 185$ är ett utfall av Y som är $\text{Po}(n_2\mu_2)$. Vi vill testa $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$ på nivå $\alpha = 0.01$.

Vi skattar $n_1\mu_1$ med 54 vilket är större än 15 och alltså är $\text{Po}(n_1\mu_1) \approx N(n_1\mu_1, \sqrt{n_1\mu_1})$. På samma sätt skattas $n_2\mu_2$ med 185 vilket är större än 15 och således är $\text{Po}(n_2\mu_2) \approx N(n_2\mu_2, \sqrt{n_2\mu_2})$.

Vi skattar μ_1 med x/n_1 och μ_2 med y/n_2 . Då gäller att θ^* som är lika med $X/n_1 - Y/n_2$ är approximativt normalfördelade eftersom en linjärkombination av två oberoende Normalfördelade stokastiska variabler är Normalfördelad. Vi kan alltså använda oss av §12.3 i formelsamlingen för att få ett konfidensintervall med approximativ konfidensgrad.

$$I_{\theta_{obs}^*} = \theta_{obs}^* \pm D_{obs}^* \cdot \lambda_{\frac{\alpha}{2}}$$

$$D = \sqrt{V(\theta^*)}$$

$$V(\theta^*) = V\left(\frac{X}{n_1} - \frac{Y}{n_2}\right) = \frac{V(X)}{n_1^2} + \frac{V(Y)}{n_2^2} = \frac{n_1\mu_1}{n_1^2} + \frac{n_2\mu_2}{n_2^2} = \frac{\mu_1}{n_1} + \frac{\mu_2}{n_2}$$

$$\Rightarrow D = \sqrt{V(\theta^*)} = \sqrt{\frac{\mu_1}{n_1} + \frac{\mu_2}{n_2}}$$

$$\Rightarrow D_{obs}^* = \sqrt{V(\theta^*)_{obs}^*} = \sqrt{\frac{\mu_{1_{obs}}^*}{n_1} + \frac{\mu_{2_{obs}}^*}{n_2}} = \sqrt{\frac{x}{n_1^2} + \frac{y}{n_2^2}}$$

$$\Rightarrow I_{\theta_{obs}^*} = \theta_{obs}^* \pm D_{obs}^* \cdot \lambda_{\frac{\alpha}{2}} = \frac{x}{n_1} - \frac{y}{n_2} \pm \sqrt{\frac{x}{n_1^2} + \frac{y}{n_2^2}} \lambda_{0.005} =$$

$$= \frac{54}{20} - \frac{185}{50} \pm \sqrt{\frac{54}{20^2} + \frac{185}{50^2}} \cdot 2.5758 = -1 \pm 1.18 = (-2.18, 0.18)$$

$0 \in I_{\mu_1 - \mu_2}$ varför hypotesen H_0 inte kan förkastas på risknivån 1%.