Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/SF1914/SF1915/SF1916 SANNOLIKHETSTEORI OCH STATISTIK, MÅNDAG 21 OKTOBER 2024 KL 8.00–13.00.

*Examinator SF1914/SF1915/SF1916*: Mykola Shykula, 08-790 66 44
*Examinator SF1912*: Celia Garcia Pareja, 072-872 34 06

*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 flerval uppgifterna 1–12, där i varje uppgift *endast ett svarsalternativ är korrekt*. På Del I ska endast svar i form av det korrekta svarsalternativet anges. 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 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"). Bonusen gäller för den här tentamen och vid omtentamen i december 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. På denna del ska resonemang och uträkningar vara så utförliga och väl motiverade att de är lätta att följa och förstå. Införda beteckningar ska förklaras och definieras samt numeriska svar ska anges med minst tre värdesiffrors noggrannhet. 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. Studenter som är godkända på den andra datorlaborationen får dessutom tre bonuspoäng på del II. Dessa bonuspoäng gäller för den här tentamen och vid omtentamen i december 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

En undersökning görs för att fastställa antalet hushåll som har elektriska apparater i en viss stad. Det visar sig att 75% har radioapparater, 65% har strykjärn, 55% har elektriska brödrostar. Vidare har 50% både radioapparater och strykjärn, 40% har både radioapparater och brödrostar och 30% har både strykjärn och brödrostar. Vi vet också att 95% hushåll har minst en av dessa apparater. Beräkna sannolikheten att ett hushåll har alla tre dessa apparater?

A: 0.80

B: 0.25

C: 0.20

D: 0.05

### Uppgift 2

En stokastisk variabel $X$ har täthetsfunktionen

$$f_X(x) = \begin{cases} \dfrac{x^3}{2a}, & 0 \leq x \leq a, \\ 0, & \text{annars,} \end{cases}$$

där $a > 0$ är en okänd konstant. Bestäm väntevärdet $E(X^4)$.

A: 1

B: 2

C: 4

D: 8

### Uppgift 3

För de två diskreta stokastiska variablerna $X$ och $Y$ gäller att

$$P(X = 0, Y = 0) = 0.10, \quad P(X = 0, Y = 1) = 0.20, \quad P(X = 1, Y = 0) = 0.30,$$

dvs $X$ och $Y$ kan endast anta värden 0 eller 1. Bestäm kovariansen $C(X, Y)$.

A: 0.42

B: 0.4

C: 0.0024

D: -0.02

### Uppgift 4

Flygplatstransporten från terminalen till hyrbils- och långtidsparkeringscentralen är tänkt att anlända högst var åttonde minut. Väntetiderna för transporten är kända för att följa en likformig fördelning. Beräkna sannolikheten att en passagerare väntar på flygplatstransporten i mer än sju minuter givet att hen redan har väntat minst fyra minuter?

A: 0.125

B: 0.250

C: 0.375

D: 0.500

## Uppgift 5

En engelsktalande turist besöker en stad i ett land där 30% av befolkningen talar engelska. Ibland måste man fråga någon om vägen. Därför, under en dag, frågar turisten fyra på måfå valda (dvs oberoende av varandra) lokala människor vägen. Vad är sannolikheten att minst två av dem kommer att kunna engelska?

A: ca 0.92

B: ca 0.26

C: ca 0.35

D: ca 0.08

## Uppgift 6

Låt $X$ vara en stokastisk variabel som beskriver poängen på den matematiska delen av Standard Achievement Test (SAT) och låt $Y$ vara en stokastisk variabel som beskriver poängen på den verbala delen av SAT. Antag att $X$ är normalfördelad med väntevärde $\mu_X = 529$ och standardavvikelse $\sigma_X = 75.7$, och $Y$ också är normalfördelad med väntevärde $\mu_Y = 474$ och standardavvikelse $\sigma_Y = 79.8$. Antag också att $X$ och $Y$ är oberoende när $X$ och $Y$ är resultat från olika personer. Vi väljer ut två elever slumpmässigt. Vad är sannolikheten att en av elevernas mattepoäng är mindre än den andra elevens verbala poäng?

A: 0.3085

B: 0.3632

C: 0.6368

D: 0.6915

## Uppgift 7

Låt $x = 95$ vara ett utfall av en stokastisk varaibel $X \in Bin(n, p)$, där $n = 100$ och $p$ är okänd. Bestäm medelfelet för skattningen $p^* = X/n$.

A: 0.0218

B: $2.256 \cdot 10^{-3}$

C: $4.75 \cdot 10^{-4}$

D: 0.0475

### Uppgift 8

Antag att vi gör två stickprov $x_1, x_2, \ldots, x_n$ och $y_1, y_2, \ldots, y_m$ från två oberoende normalfördelade populationer $N(\mu_1, \sigma_1)$ respektive $N(\mu_2, \sigma_2)$. Stickprovet $x_1, \ldots, x_n$ gav medelvärdet $\overline{x}$ och standardavvikelsen $s_x$, medan stickprovet $y_1, \ldots, y_m$ gav medelvärdet $\overline{y}$ och standardavvikelsen $s_y$. Antag vidare att i populationen $\sigma_1 = \sigma_2$ och att de är okända. Vi testar $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$, på signifikans nivån 5%. Som testvariabel används: $t = \frac{\overline{x}-\overline{y}}{s\sqrt{\frac{1}{n}+\frac{1}{m}}}$, där $s^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2}$. Om stickprovsstorlekarna är $n = 13$ och $m = 13$, så förkastas $H_0$ om

A: $t < -1.71$

B: $|t| > 1.71$

C: $t > 1.71$

D: $t < -2.06$

E: $|t| > 2.06$

F: $t > 2.06$

### Uppgift 9

Antag att $x_1, \ldots, x_n$ utgör ett stickprov från en normalfördelning $N(\mu, \sigma)$, där standardavvikelse $\sigma$ är *okänd* och $n = 10$. Emma önskar testa nollhypotesen $H_0 : \mu = 5$ mot $H_1 : \mu < 5$ med hjälp av ett lämligt konfidensintervall för $\mu$. Låt

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

Vilket av nedanstående konfidensintervall för $\mu$ kan användas för att genomföra hypotestestet ovan på signifikansnivå $\alpha$?

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

B: $I_\mu = \left(\bar{x} - t_{\alpha/2}(n-1)\,\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2}(n-1)\,\frac{s}{\sqrt{n}}\right)$

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

D: $I_\mu = \left(\bar{x} - t_{\alpha}(n-1)\,\frac{s}{\sqrt{n}},\ +\infty\right)$

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

F: $I_\mu = \left(-\infty,\ \bar{x} + t_{\alpha}(n-1)\,\frac{s}{\sqrt{n}}\right)$

## Uppgift 10

Låt $X \in Po(\mu)$. Vi testar nollhypotesen $H_0 : \mu = 4$ mot mothypotesen $H_1 : \mu > 4$. Vid ett försök erhålls obervationen $x = 7$. Bestäm testets $P$-värde.

A: ca 0.05

B: ca 0.11

C: ca 0.22

D: ca 0.95

## Uppgift 11

Man vill undersöka om följande data kommer från en binomialfördelning med parametern $n = 3$ och något värde på parametern $p$, t ex $p = 0.45$, dvs. man testar nollhypotesen $H_0$ : data kommer från $Bin(3, 0.45)$ mot $H_1$ : data kommer *inte* från $Bin(3, 0.45)$.

| | 0 | 1 | 2 | 3 | Totalt |
|---|---|---|---|---|---|
| antal observationer | 11 | 52 | 35 | 2 | 100 |
| förväntat antal observationer under $H_0$ | 17 | 41 | 33 | 9 | 100 |

Givet data ovan blir teststorheten $Q = 10.588$. Vilket av följande påståenden är sant?

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

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

Man har oberoende obervationer $x_1, x_2, \ldots, x_n$ från en fördelning med täthetsfunktionen

$$f(x) = \begin{cases} \lambda^2\, x\, e^{-\lambda\, x}, & \text{för } x > 0, \\ 0, & \text{annars,} \end{cases}$$

där $\lambda > 0$. Bestäm MK-skattningen av parametern $\lambda$.

A: $2\,\lambda \sum_{i=1}^{n} x_i$

B: $\dfrac{2\sum_{i=1}^{n} x_i}{n}$

C: $\dfrac{2n}{\sum_{i=1}^{n} x_i}$

D: $\dfrac{4n}{\sum_{i=1}^{n} x_i}$

**Var god vänd!**

# Del II

### Uppgift 13

Antag att den förväntade väntetiden för en kund när kunden ringer ett försäkringsbolag är 5 (minuter). Vi antar att väntetiderna kan modelleras med hjälp av en exponentialfördelning. Det vill säga om $X$ är väntetiden för en på måfå vald kund, så följer att den stokastiska variabeln $X$ är exponentialfördelad med väntevärdet 5 (minuter). Tre kunder ringer oberoende av varandra till försäkringsbolaget och väntar tills de blir inkopplade. Beräkna sannolikheten att den längsta väntetiden för dessa tre kunder är mindre än 4 minuter. (10 p)

### Uppgift 14

I ett laboratorieexperiment studerar forskare blodnivåerna av cykliskt adenosinmonofosfat (cAMP) i 4 grodor. Varje groda utsätts för två kemiska ämnen $C_A$ och $C_B$, varefter blodprov tas och mängden cAMP (i pmol/enhet) mäts. De erhållna resultaten sammanfattas i följande tabell:

| | cAMP (pmol/enhet) | |
|---|---|---|
| Groda | $C_A$ | $C_B$ |
| 1 | 6.01 | 5.23 |
| 2 | 2.28 | 1.21 |
| 3 | 1.51 | 1.40 |
| 4 | 2.12 | 1.38 |

Förutsatt att mängden cAMP i blod är normalfördelad, avgör med ett lämpligt hypotestest på en signifikansnivå $\alpha = 0.05$ om det föreligger någon skillnad mellan effekten av de kemiska ämnena $C_A$ och $C_B$ i mängden cAMP i blod. Ange tydligt noll- och alternativ hypotes, slutsatsen av testet och tolka resultatet (det vill säga avgör om det finns ett signifikant bevis för att kemikalierna $C_A$ och $C_B$ har olika effekter i genomsnitt). (10 p)

### Uppgift 15

Täthetsfunktionen för $Pareto(\alpha, \beta)$-fördelningen ges av

$$f(x) = \beta \frac{\alpha^\beta}{x^{\beta+1}}, \quad x \geq \alpha, \quad \alpha, \beta > 0.$$

Antag att vi har $n$ observationer $x_1, \ldots, x_n$ som är utfall av oberoende stokastiska variabler med ovanstående fördelning, där $\alpha$ och $\beta$ är positiva men okända parametrar. Härled ML-skattningarna av parametrarna $\alpha$ och $\beta$. (10 p)

**Uppgift 16**

Ett sjukhus ska placeras ut i ett kvadratiskt samhälle. Samhället representeras av $K = \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1,\ 0 \le y \le 1\}$, där punkten $(0, 0)$ motsvarar samhällets sydvästra hörn. Gatusystemet är beskaffat så att man endast kan färdas *parallellt med koordinataxlarna*. På grund av faktorer som t ex befolkningstäthet antas antalet olyckor öka linjärt i både $x$ och $y$-led. Platsen $(X, Y)$ för en olycka kan därför beskrivas med en tvådimensionell kontinuerlig stokastisk fördelning vars täthetsfunktionen är

$$f_{X,Y}(x, y) = \begin{cases} 4xy, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{för övrigt.} \end{cases}$$

(a) Vad är den förväntade utryckningssträckan för ambulansen (dvs sträckan från olyckplatsen till sjukhuset), om sjukhuset placeras i samhällets sydvästra hörn (dvs i punkten $(0, 0)$)? (3 p)

(b) Bestäm den placering av sjukhuset, dvs punkten $(a, b) \in K$, som minimerar den förväntade utryckningssträckan. (7 p)

**Lycka till!**

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1912/SF1914/SF1915/SF1916
SANNOLIKHETSTEORI OCH STATISTIK,
MÅNDAG 21 OKTOBER 2024 KL 8.00–13.00.

## Del I, Svar.

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

**Del I, Lösningsförslag.**

### Uppgift 1

Minst en av dessa aparater (R=Radio,S=Stryk,B=Bröd): $P(R \cup S \cup B) = 0.95$
Vi vet att:

$$P(R \cup S \cup B) = P(R) + P(S) + P(B) - P(R \cap S) - P(R \cap B) - P(S \cap B) + P(R \cap S \cap B),$$

eller, efter att vi stoppar in allt som vi känner till,

$$0.95 = 0.75 + 0.65 + 0.55 - 0.5 - 0.4 - 0.3 + P(R \cap S \cap B),$$

vilket ger $P(R \cap S \cap B) = 0.95 - 0.75 - 0.65 - 0.55 + 0.5 + 0.4 + 0.3 = 0.2$

Svaret är $C$.

### Uppgift 2

Först bestämmer vi $a$. För täthetsfunktionen $f_X(x)$ gäller: $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Vi har därför:

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_0^a \frac{x^3}{2a}\,dx = \left[\frac{x^4}{8a}\right]_{x=0}^{x=a} = \frac{a^4}{8a} = \frac{a^3}{8},$$

som ger att $a = 2$. Vidare, har vi:

$$E(X^4) = \int_{-\infty}^{\infty} x^4 f_X(x)dx = \int_0^2 x^4 \frac{x^3}{4}\,dx = \frac{1}{4}\int_0^2 x^7 dx = \frac{1}{4}\left[\frac{x^8}{8}\right]_{x=0}^{x=2} = \frac{2^8}{2^2 \cdot 2^3} = 2^3 = 8$$

Svaret är $D$.

### Uppgift 3

Eftersom hela sannolikhetsmassa är lika med 1, och både $X$ och $Y$ kan endast anta värden 0 eller 1, så följer det att

$$P(X = 1, Y = 1) = 1 - (0.1 + 0.2 + 0.3) = 0.4$$

Videra, den marginella fördelningen för $X$: $P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = 0.1 + 0.2 = 0.3$ och $P(X = 1) = 0.3 + 0.4 = 0.7$, och därmed väntevärdet

$$E(X) = (0)0.3 + (1)0.7 = 0.7$$

På samma sätt, den marginella fördelningen för $Y$: $P(Y = 0) = 0.1 + 0.3 = 0.4$ och $P(Y = 1) = 0.2 + 0.4 = 0.6$, och därmed väntevärdet

$$E(Y) = (0)0.4 + (1)0.6 = 0.6$$

Slutligen, med hänsyn till informationen att $E(X) = 0.7$ och $E(Y) = 0.6$, har vi:

$$C(X,Y) = E(XY) - E(X)E(Y) = (0)(0)0.1 + (0)(1)0.2 + (1)(0)0.3 + (1)(1)0.4 - E(X)E(Y) =$$
$$= 0.4 - E(X)E(Y) = 0.4 - (0.7)(0.6) = -0.02$$

Svaret är $D$.

## Uppgift 4

Låt $X$ vara väntetiden (i min) för flygplanstransporten. Vi har då att $X \in U(0,8)$, dvs

$$f_X(x) = \begin{cases} \frac{1}{8}, & \text{om } 0 < x < 8 \\ 0, & \text{annars.} \end{cases}$$

Därmed, har vi:

$$P(X > 7 | X \geq 4) = P(X \geq 7 | X \geq 4) = \frac{P(\{X \geq 7\} \cap \{X \geq 4\})}{P(X \geq 4)} = \frac{P(X \geq 7)}{P(X \geq 4)}$$
$$= \frac{\int_7^8 \frac{1}{8} dx}{\int_4^8 \frac{1}{8} dx} = \frac{\frac{1}{8}x\Big|_7^8}{\frac{1}{8}x\Big|_4^8} = \frac{\frac{1}{8}}{\frac{4}{8}} = \frac{1}{4} = 0.25$$

Svaret är $B$.

## Uppgift 5

Låt $X$ vara en s.v. som anger antal engelsktalande av fyra på måfå valda lokala människor. Då följer s.v. $X$ en binomialfördelning med parametrarna 4 och 0.3, dvs $X \in Bin(4, 0.3)$.
Vi söker: $P(X \geq 2) = 1 - P(X \leq 1)$. Enligt Tabell 6 för $X \in Bin(4, 0.3)$, har vi:

$$P(X \leq 1) = 0.65170$$

Slutligen,

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.65170 \approx 0.35$$

Svaret är $C$.

## Uppgift 6

Från lydelsen, vet vi att $X \in N(\mu_X, \sigma_X)$, där $\mu_X = 529$, $\sigma_X = 75.7$, och $Y \in N(\mu_Y, \sigma_Y)$, där $\mu_Y = 474$, $\sigma_Y = 79.8$
Vi har: $P(X < Y) = P(X - Y < 0)$. Videra, vi har också:

$$E(X - Y) = E(X) - E(Y) = 529 - 474 = 55$$

och

$$V(X - Y) = 1^2 V(X) + (-1)^2 V(Y) = 75.7^2 + 79.8^2 = 12098.53$$

Nu, eftersom en linjär kombination av normalfördelade stokastiska variabler är också en normalfördelad stokastisk variabel, vi har att stokastisk variabel $X - Y$ är normalfördelad med väntevärdet $\mu = E(X - Y) = 55$ och standardavvikelsen $\sigma = D(X - Y) = \sqrt{V(X - Y)} = \sqrt{12098.53} = 109.99$, dvs

$$X - Y \in N(\mu, \sigma),$$

där $\mu = 55$ och $\sigma = 109.99$
Därmed, får vi mha $z$-transformationen:

$$P(X < Y) = P(X - Y < 0) = P\left(Z < \frac{0 - \mu}{\sigma}\right) = P\left(Z < \frac{0 - 55}{109.99}\right) = P\left(Z < -\frac{1}{2}\right) =$$

$$= 1 - P\left(Z < \frac{1}{2}\right) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085,$$

där $Z \in N(0, 1)$, och vi använde Tabell 1 för att ta redo på $\Phi(0.5)$.

Svaret är $A$.

## Uppgift 7

Först har vi variansen för $p^* = X/n$:

$$V(p^*) = V(X/n) = V(X)/n^2 = (np(1 - p))/n^2 = p(1 - p)/n.$$

Därför blir standardavvikelsen för skattningen $p^*$

$$D(p^*) = \sqrt{V(p^*)} = \sqrt{p(1 - p)/n},$$

och därmed medelfelet för $p^*$ blir:

$$d(p^*) = \Big(D(p^*)\Big)^*_{obs} = \left(\sqrt{p(1 - p)/n}\right)^*_{obs} = \sqrt{p^*_{obs}(1 - p^*_{obs})/n} = \sqrt{(x/n)(1 - x/n)/n} =$$

$$= \sqrt{x(n - x)/n^3} = \sqrt{95(100 - 95)/100^3} \approx 0.021794$$

Svaret är $A$.

## Uppgift 8

Eftersom $H_1$ är tvåsidig och $\sigma_1 = \sigma_2$ är okända, så förkastas $H_0$ om

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

där $n = 13$, $m = 13$, och $\alpha = 0.05$. Dvs, $t_{\alpha/2}(n + m - 2) = t_{0.025}(24) = 2.06$, enligt Tabell 3.

Svaret är $E$.

## Uppgift 9

Eftersom $H_1 : \mu < 5$ är ensidig och pekar mot $-\infty$ och $\sigma$ är okänd dvs måste skattas med $s$, följande isf (ensidig) uppåt begränsad k.i. kan användas för att testa $H_0 : \mu = 5$ mot $H_1 : \mu < 5$:

$$I_\mu = \left(-\infty,\, \bar{x} + t_\alpha(n - 1)\,\frac{s}{\sqrt{n}}\right).$$

Svaret är $F$.

**Uppgift 10**

Vi har enl Tabell 5 under $H_0$:

$$P - value = P(X \geq 7 | X \in Po(4)) = 1 - P(X \leq 6 | X \in Po(4)) = 1 - 0.88933 \approx 0.11$$

Svaret är $B$.

**Uppgift 11**

Eftersom alla förväntade antal är större än 5, behöver vi inte slå ihop grupperna. Vidare, eftersom parameter $p = 0.45$ är känd under $H_0$, så behöver vi inte skatta någon parameter. Beslutsregeln kommer att vara: förkasta $H_0$ på signifikansnivån (risknivån) $\alpha$ om

$$Q > \chi^2_\alpha(r-1),$$

där $r$ är antal grupper (i denna uppgift är antal grupper $r = 4$), och $Q = 10.588$.

Därmed, enligt Tabell 4, kan vi förkasta $H_0$ på risknivån 5%, eftersom

$$Q = 10.588 > 7.81 = \chi^2_{0.05}(3),$$

och kan inte förkasta $H_0$ på risknivån 1%, eftersom

$$Q = 10.588 < 11.3 = \chi^2_{0.01}(3).$$

Svaret är $D$.

**Uppgift 12**

Låt $x_1, x_2, \ldots, x_n$ vara observationer av s.v. $X_1, X_2, \ldots, X_n$ med täthetsfunktionen

$$f(x) = \begin{cases} \lambda^2 \, x \, e^{-\lambda\, x}, & \text{för } x > 0, \\ 0, & \text{annars,} \end{cases}$$

där $\lambda > 0$. För att bestämma MK-skattningen, behöver vi först beräkna väntevärdet $E(X_i)$ som blir samma för alla $1 \leq i \leq n$. Vi har:

$$E(X_i) = \int_0^\infty x^2 \lambda^2 e^{-\lambda x} dx = \lambda \int_0^\infty x^2 \big(\lambda e^{-\lambda x}\big) dx,$$

där i sin tur för en s.v. $X \in Exp(\lambda)$ (som är oberoende av $X_1, X_2, \ldots, X_n$) gäller enl definitionen av variansen att

$$\int_0^\infty x^2 \big(\lambda e^{-\lambda x}\big) dx = E(X^2) = V(X) + \big(E(X)\big)^2 = \frac{1}{\lambda^2} + \Big(\frac{1}{\lambda}\Big)^2 = \frac{2}{\lambda^2}.$$

Därför har vi nu:

$$E(X_i) = \lambda \int_0^\infty x^2 \big(\lambda e^{-\lambda x}\big) dx = \lambda \, \frac{2}{\lambda^2} = \frac{2}{\lambda}.$$

MK-skattningen nu. Först sätter vi upp $Q$-funktionen, och vi har:

$$Q = Q(\lambda) = \sum_{i=1}^{n} \big(x_i - E(X_i)\big)^2 = \sum_{i=1}^{n} \Big(x_i - \frac{2}{\lambda}\Big)^2.$$

Nu deriverar vi $Q$ map $\lambda$, sätter till 0, och löser sedan ekvationen för $\lambda$. Vi får:

$$\frac{dQ(\lambda)}{d\lambda} = \sum_{i=1}^{n} 2\Big(x_i - \frac{2}{\lambda}\Big)(-1)(-1)\frac{2}{\lambda^2} = 0,$$

som förblir

$$\sum_{i=1}^{n} \Big(x_i - \frac{2}{\lambda}\Big) = 0 \quad \text{eller} \quad \sum_{i=1}^{n} x_i = \frac{2n}{\lambda},$$

som ger till slut lösningen

$$\lambda^* = \lambda^*_{obs,MK} = \frac{2n}{\sum_{i=1}^{n} x_i}.$$

Svaret är $C$.

**Del II, Lösningsförslag.**

## Uppgift 13

Vi har att $X_i \in Exp(\lambda)$, är väntetiden för en kund när kunden ringer ett försäkringsbolag, med den förväntade väntetiden lika med 5 minuter, dvs $E(X_i) = 5 = 1/\lambda \Rightarrow \lambda = 1/5 = 0.2$, $i = 1, 2, 3$. Också, eftersom $X_i \in Exp(\lambda)$ vet vi att:

$$f_{X_i}(t) = \begin{cases} \lambda \exp\{-\lambda t\}, \text{ om } t \geq 0, \\ 0, \text{ annars.} \end{cases}$$

Därför, ty fördelningsfunktionen $F_{X_i}(u) = P(X_i \leq u) = \int_0^u \lambda \exp\{-\lambda t\} dt = \Big[ - \exp\{-\lambda t\} \Big]_{t=0}^{t=u}$, vi har

$$F_{X_i}(u) = P(X_i \leq u) = \begin{cases} 1 - \exp\{-\lambda u\}, \text{ om } \geq 0, \\ 0, \text{ annars,} \end{cases}$$

där $\lambda = 0.2$ och $i = 1, 2, 3$

Vidare, låt $X_1$, $X_2$ och $X_3$ vara väntetider för de tre kunderna 1, 2 och 3, respektive, som ringer oberoende av varandra. Och, låt s.v. $Y = \max(X_1, X_2, X_3)$. Fördelningsfunktionen för $Y$ (dvs det vi söker egentligen) blir:

$$\begin{aligned} P(Y \leq y) &= P(\max(X_1, X_2, X_3) \leq y) = P(X_1 \leq y, X_2 \leq y, X_3 \leq y) = \\ &= P(X_1 \leq y)P(X_2 \leq y)P(X_3 \leq y) = \big(P(X_1 \leq y)\big)^3 = \Big(1 - e^{-(0.2)y}\Big)^3, \end{aligned}$$

ty om maximum är mindre än $y$, så måste samtliga $X_1$, $X_2$ och $X_3$ också vara mindre än $y$. Också, har använt här ovanför att $X_1$, $X_2$ och $X_3$ är oberoende; och att $X_1$, $X_2$ och $X_3$ har samma fördelning (dvs $X_i \in Exp(0.2)$, $i = 1, 2, 3$). Slutligen, eftersom fördelningen är kontinuerlig:

$$P(Y < 4) = P(Y \leq 4) = \Big(1 - e^{-(0.2)4}\Big)^3 \approx 0.167$$

**Svar:** 0.167

## Uppgift 14

Stickprov i par. Signifikansnivån $\alpha = 0.05$. Tvåsidig alternativ hypotes $H_1 : \mu_{i,C_A} \neq \mu_{i,C_B}$, $1 \leq i \leq n$, $n = 4$.

Låt nu $X = X_{C_A} \in N(\mu + \Delta, \sigma_X)$ och $Y = Y_{C_B} \in N(\mu, \sigma_Y)$. Och sedan $X - Y = Z \in N(\Delta, \sigma)$, där från datamaterilet $z_1 = x_1 - y_1 = 6.01 - 5.23 = 0.78, \ldots, z_4 = x_4 - y_4 = 2.12 - 1.38 = 0.74$, vilket ger oss: $\Delta^*_{obs} = \overline{z} = 0.675$ och $\sigma^*_{z,obs} = \sqrt{s_z^2} = 0.404$.

Vi testar

$$H_0 : \Delta = 0$$

mot

$$H_1 : \Delta \neq 0,$$

och det kritiska värdet $t_{\alpha/2}(n-1) = t_{0.05/2}(3) = 3.18$ (enl Tabell 3). Vidare, den observerade testvariabeln är:

$$t_{obs} = \frac{\Delta^*_{obs} - 0}{\sigma^*_{z,obs}/\sqrt{n}} = \frac{\bar{z} - 0}{\sqrt{s_z^2}/\sqrt{n}} = \frac{0.675 - 0}{0.404/\sqrt{4}} = 3.4$$

Och beslutsregeln är: förkasta $H_0$ om $|t_{obs}| > t_{\alpha/2}(n-1)$. Eftersom, $|t_{obs}| = 3.4 > 3.18 = t_{0.05/2}(3)$, kan vi förkasta våran nollhypotes på signifikansnivån $\alpha = 0.05$.

**Slutsats:** på signifikansnivån $\alpha = 0.05$ föreligger det en signifikant skillnad mellan effekten av de kemiska ämnena $C_A$ och $C_B$ i mängden cAMP i blod.

## Uppgift 15

Eftersom $x_i$:na antas vara oberoende observationer, blir likelihood-funktionen

$$L(\alpha,\beta) = L(x_1, x_2, \ldots, x_n; \alpha, \beta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \beta \frac{\alpha^\beta}{x_i^{\beta+1}} = \beta^n \alpha^{n\beta} \prod_{i=1}^{n} x_i^{-(\beta+1)}.$$

där $x_i \geq \alpha$ för alla $i$, och $\alpha, \beta > 0$. Vi vill maximera funktionen $L(\alpha,\beta)$ m.a.p. $\alpha$ och $\beta$. Vi logaritmerar först $L(\alpha,\beta)$ för att få log-likelihood funktionen $ln(L(\alpha,\beta))$, vi har:

$$ln(L(\alpha,\beta)) = n\,ln(\beta) + n\beta\,ln(\alpha) - (\beta+1)\sum_{i=1}^{n} ln(x_i).$$

Nu deriverar vi funktionen $ln(L(\alpha,\beta))$ m.a.p. på $\alpha$ resp $\beta$ och sätter till noll. Vi får:

$$\begin{cases} \frac{\partial ln(L(\alpha,\beta))}{\partial\alpha} = \frac{n\beta}{\alpha} = 0, \\ \frac{\partial ln(L(\alpha,\beta))}{\partial\beta} = \frac{n}{\beta} + n\,ln(\alpha) - \sum_{i=1}^{n} ln(x_i) = 0. \end{cases}$$

Från den första ekvationen - får vi tyvärr inte så mycket information. Från den andra ekvationen, ser vi att det går lätt att lösa ut $\beta$ om vi nu vet vad $\alpha$ är, vi får då:

$$\beta^*_{obs,ML} = \frac{n}{\sum_{i=1}^{n} ln(x_i) - n\,ln(\alpha^*_{obs,ML})}.$$

Alltså, vi måste ta redo på vad $\alpha^*_{obs,ML}$ är.

För att erhålla även $\alpha^*_{obs,ML}$, resonerar vi oss tillbaka till funktionen $L(\alpha,\beta)$ som ska maximeras m.a.p. $\alpha$. Men, om vi tittar noggrannt på $L(\alpha,\beta)$ ser vi att - ett högre $\alpha$ ger alltid ett högre värde på $L(\alpha,\beta)$ (dvs det finns inget maximum). Samtidigt kommer vi ihåg från definitionen att

$$\alpha \leq x_i,$$

för alla $i = 1, \ldots, n$. Således, ML-skattningen för parametern $\alpha$ blir

$$\alpha^*_{obs,ML} = \min_{1 \leq i \leq n} \{x_i\}.$$

**Uppgift 16**

(a) Vi har:

$$E(X+Y) = \int_0^1 dx \int_0^1 (x+y)4xydy =$$

$$= 4\int_0^1 x^2dx \int_0^1 ydy + 4\int_0^1 xdx \int_0^1 y^2dy = 4\big(1/3\big)\big(1/2\big) + 4\big(1/2\big)\big(1/3\big) = 4/3.$$

**Svar:** $4/3$

(b) Vi söker punkten $(a,b) \in K$ som minimerar funktionen:

$$E\big(|X-a| + |Y-b|\big) = E\big(|X-a|\big) + E\big(|Y-b|\big).$$

Eftersom allt är symmetriskt, så räcker det att bestämma den ena termen i summan ovan, tex den andra. Täthetsfunktionen för s.v. $Y$ blir:

$$f_Y(y) = \int_0^1 4xydx = 2y, \ \ 0 \le y \le 1.$$

Därför, har vi:

$$E\big(|Y-b|\big) = \int_0^1 2y|y-b|dy =$$

$$= \int_0^b 2y(b-y)dy + \int_b^1 2y(y-b)dy = \frac{2}{3}b^3 - b + \frac{2}{3}.$$

Alltså, tillsammans, har vi att

$$E\big(|X-a| + |Y-b|\big) = E\big(|X-a|\big) + E\big(|Y-b|\big) = \frac{2}{3}a^3 - a + \frac{2}{3} + \frac{2}{3}b^3 - b + \frac{2}{3} = g(a,b).$$

Vi får då, om vi deriverar den senaste funktionen $g(a,b)$ (m.a.p. på $a$ respektive $b$) och sätter till noll, att lösningen av dom två ekvationerna är punkten $\left(\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right)$.
Detta blir också den placering som *minimerar* funktionen $E\big(|X-a|+|Y-b|\big)$, dvs den förväntade utrykningssträckan, eftersom andra derivata för $g(a,b)$ blir $4a$ respektive $4b$, som både är strikt positiva i $\sqrt{\frac{1}{2}}$.

**Svar:** $\left(\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right)$