

*Hjälpmedel: Räknedosa, formel- och tabellsamling och engelsk-svensk ordlista för kursen 1MS321. Goda resonemang och motiveringar vägs in vid poängsättning. För betygen 3, 4, resp 5 krävs normalt minst 18, 25, resp 32 poäng.*

**Fråga 1:** Låt  $X$  vara en slumpvariabel med täthetsfunktion  $f_X(x) = 3x^2$  när  $0 \leq x \leq 1$  och  $f_X(x) = 0$  annars.

a) Beräkna  $E[X]$  och  $V[X]$ . (2p)

b) Bestäm  $P(X = \frac{1}{2})$  och  $P(X \geq \frac{1}{2})$ . (2p)

c) Beräkna  $C(X, X^2)$ , det vill säga kovariansen mellan slumpvariablerna  $X$  och  $X^2$ . (2p)

**Fråga 2:** Låt  $X$  vara en slumpvariabel med täthetsfunktion

$$f(x) = \frac{2}{9}x \quad \text{när } 0 \leq x \leq 3$$

och  $f(x) = 0$  annars. Simulera tre pseudoslumptal från sannolikhetsfördelningen av  $X$  med hjälp av följande pseudoslumptal från  $\text{Re}[0, 1]$ :

$$u_1 = 0.5644 \quad u_2 = 0.2375 \quad u_3 = 0.9252$$

(5p)

**Fråga 3:** Låt  $X_1 \sim N(\mu, 4)$  och  $X_2 \sim N(2\mu, 9)$  vara två oberoende slumpvariabler, där  $\mu$  är en okänd parameter.

a) Visa att  $T_1 = \frac{1}{2}X_1 + \frac{1}{4}X_2$  är en väntevärdesriktig skattning av  $\mu$ . (2p)

b) Beräkna standardavvikelsen av  $T_1$ . (1p)

c) Låt  $T_2 = 2X_1 - cX_2$ , där  $c$  är en konstant. Vad måste  $c$  vara så att  $T_2$  är en väntevärdesriktig skattning av  $\mu$ ? Givet detta  $c$ , vilken skattning är mest effektiv,  $T_1$  eller  $T_2$ ? (3p)

**Fråga 4:** Svarsfördröjningen av en server på ett företag antas vara normalfördelad med okänt väntevärde och varians. Företaget vill mäta fördröjningen och gör 20 anrop till servern. Denna mätning ger ett stickprov  $x_1, \dots, x_{20}$ , uppmätt i millisekunder, vilka antas alla vara ömsesidigt oberoende. Från detta beräknas stickprovsmedelvärdet 42ms och stickprovsstandardavvikelsen 10ms.

a) Ge ett 95% konfidensintervall för svarsfördröjningen av servern. (2p)

b) Antag istället att standardavvikelsen är känd och att den är exakt  $\sigma = 10\text{ms}$ . Hur många observationer behöver man genomföra för att få ett 95% konfidensintervall vars längd är som mest 5ms? (3p)

**Fråga 5:** I ett socialt medium är de flesta användare människor, men ungefär 5% av alla användare är bottar. Det genomsnittliga antalet dagliga interaktioner hos en mänsklig användare antas följa en normalfördelning med väntevärde 18 och standardavvikelse 5, oberoende av alla andra användare. Användare som är bottar har betydligt många mer interaktioner. Deras genomsnittliga antal dagliga interaktioner antas följa en normalfördelning med väntevärde 28 och standardavvikelse 5, oberoende av alla andra användare.

Introducera en lämplig modell, som definierar slumpvariabler och händelser för att svara på följande:

- Vad är sannolikheten att en given mänsklig användare har genomsnittligen mer än 30 dagliga interaktioner? Vad är sannolikheten att en given botanvändare har genomsnittligen mer än 30 dagliga interaktioner? (3p)
- Företaget bakom det sociala mediet överväger ett radikalt förslag för att minska antalet bottar på plattformen: Radera alla användare som har genomsnittligen mer än 30 dagliga interaktioner. Vad är sannolikheten att en användare med genomsnittligen mer än 30 dagliga interaktioner tillhör en människa? (3p)

**Fråga 6:** Låt  $X_0, X_1, X_2, \dots$  vara en Markovkedja med tillståndsrum  $E = \{0, 1, 2\}$  och övergångsmatris

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1 \\ 2/3 & 0 & * \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Markovkedjan startar antingen vid tillståndet 0 eller 2, båda med samma sannolikhet.

- Vilket värde ska det saknade värdet (\* i matrisen) vara, varför? (1p)
- Beräkna  $P(X_2 = 1)$ . (2p)
- Man kan visa att det finns en unik stationär fördelning för övergångsmatrisen  $\mathbf{P}$ . Bestäm denna stationära fördelning. (3p)

**Fråga 7:** I ett datanätverk finns två servrar, A och B, där dataintrångsförsök kan ske. Dataintrångsförsöken kan modelleras som två oberoende Poissonprocesser  $\{N_A(t) : t \geq 0\}$  och  $\{N_B(t) : t \geq 0\}$ , med tillhörande intensiteter  $\lambda_A = 0.001$  per timma (för server A) samt  $\lambda_B = 0.003$  per timma (för server B).

- Vad är sannolikheten att det inte sker något dataintrångsförsök hos server A under en dag? (2p)
- Vad är sannolikheten att det sker åtminstone tre dataintrångsförsök totalt (från både A och B tillsammans) inom tio dagar? (2p)
- Datanätverket hade exakt ett dataintrångsförsök idag. Vad är sannolikheten att detta hände vid server A? (2p)